Time-Division is Optimal for Covert Communication over Some Broadcast Channels

Vincent Y. F. Tan, Senior Member, IEEE  
Si-Hyeon Lee, Member, IEEE

Abstract—We consider a covert communication scenario where a transmitter wishes to communicate simultaneously to two legitimate receivers while ensuring that the communication is not detected by an adversary, the warden. The legitimate receivers and the adversary observe the transmission from the transmitter via a three-user discrete or Gaussian memoryless broadcast channel. We focus on the case where the “no-input” symbol is not redundant, i.e., the output distribution at the warden induced by the no-input symbol is not a mixture of the output distributions induced by other input symbols, so that the covert communication is governed by the square root law, i.e., at most $\Theta(\sqrt{n})$ bits can be transmitted over $n$ channel uses. We show that for such a setting, a simple time-division strategy achieves the optimal throughputs for a non-trivial class of broadcast channels; this is not true for communicating over broadcast channels without the covert communication constraint. Our result implies that a code that uses two separate optimal point-to-point codes each designed for the constituent channels and each used for a fraction of the time is optimal in the sense that it achieves the best constants of the $\sqrt{n}$-scaling for the throughputs. Our proof strategy combines several elements in the network information theory literature, including concave envelope representations of the capacity regions of broadcast channels and ElGamal’s outer bound for more capable broadcast channels.

Index Terms—Covert communication, Low probability of detection, Broadcast channels, Time-division, Concave envelopes

I. INTRODUCTION

There has been a recent surge of research interest in reliable communications in the presence of an adversary, or a warden, who must be kept incognizant of the presence of communication between the transmitters and receivers. The lack of communication is modelled in discrete channels by sending a specially-designed “no-input” symbol $0 \in \mathcal{X}$ (where $\mathcal{X}$, a finite set, is the input alphabet of the discrete channel); in Gaussian channels, it is also modelled as sending $0 \in \mathbb{R}$. This line of research, known synonymously as covert communications, communication with low probability of detection (LPD) [2–4], deniability [5], [6], or undetectable communication [7], seeks to establish fundamental limits on the throughputs to communicate to the legitimate receiver(s) while ensuring that the signals observed by the warden are statistically close to the signals if communication were not present. It was shown by Bash et al. [2] that in the point-to-point setting, if the legitimate user’s channel and the adversary’s channel are perfectly known, the number of bits that can be reliably and covertly transmitted over $n$ channel uses scales at most as $\Theta(\sqrt{n})$ for additive white Gaussian noise (AWGN) channels. This is colloquially known as the square root law (SRL). For discrete memoryless channels, the covert communication is also governed by the SRL if the no-input symbol is not redundant, i.e., the output distribution at the warden induced by the no-input symbol is not a mixture of the output distributions induced by other input symbols. Recently, the optimal pre-constant in the $\Theta(\sqrt{n})$ term has also been established by Bloch [8] and Wang, Wornell and Zheng [9].

In this paper, we extend the above model and results to a multi-user scenario [10] in which there is one transmitter, two legitimate receivers and, as usual, one warden. See Fig. 1. We are interested in communicating reliably and simultaneously to the two receivers over the same medium while ensuring that the warden remains incognizant of the presence of any communication. We call our model a two-user discrete memoryless broadcast channel (BC) with a warden. This communication model mimics the scenario of a military general delivering commands to her/his multiple subordinates while, at the same time, ensuring that the probability of the communication being detected by a furtive enemy, the warden, is vanishingly small. Note that the enemy is not interested in the precise commands per se but on whether or not communication between the general and her/his subordinates is actually happening in order to pre-empt a possible attack. We establish the fundamental limits for communicating in this scenario when the no-input symbol is not redundant. Somewhat surprisingly, we show that the most basic multi-user communication scheme of time-division [10] Sec. 5.2) is optimal for a class of BCs. This implies that a code designed for such BCs that uses two separate point-to-point codes, each designed for the constituent channels and each used for a fraction of the time (blocklength) is optimal in the sense that it achieves the best constants of the $\sqrt{n}$-scaling of the throughput. We would like to emphasize that time-division is not optimal for the vast majority of BCs in the absence of the covert communication constraint [10] Chapters 5 & 8]. In fact, the set of BCs for which time-division is optimal without the covert communication constraint has measure zero but with the covert communication constraint,
A. Related Work

Covert communication is related with steganography in the sense that both aim to hide information against a warden. In many instances of steganography, the SRL has been observed \([11]–[16]\). In steganography, a cover text is given to the encoder and a message is embedded into the cover text when the communication is active. The message-embedded text is called the stegotext and the warden should not be able to distinguish the stegotext from the cover text. In particular, a setup closely related to our problem is the scenario in which the stegotext is generated through a memoryless channel as in \([13]–[15]\). It is interesting to observe that for steganography, a source sequence (cover text) is given to the encoder and the encoder controls the channel (regarding the cover text as the input and the stegotext as the output), while in covert communications, the encoder generates a source sequence (codeword) for each message and then the source sequence is sent through a given communication channel. Since the given condition and the control function are swapped, the two classes of problems require different analyses. For example, the positive steganography rate shown in \([17]\) relies on the fact that the encoder can modify some part of the covert text, while such a technique is not applicable in standard communication systems unless the transmitter has access to the warden’s noise.

This paper focuses on the information-theoretic aspects of covert communications for which there has been a flurry of recent work. In particular, refined asymptotics on the fundamental limits of covert communications over memoryless channels from the second-order \([18]\) and error exponent \([19]\) perspectives have been studied. In addition, the fundamental limits of covert communications for channels with random state known at the transmitter \([20]\) and classical-quantum channels \([21]\) have also been established. In contrast, work on multi-user extensions of covert communications is relatively sparse. Of note are the works by Arumugam and Bloch \([22]\), \([23]\). In \([22]\), the authors derived the fundamental limits of covert communications over a multiple-access channel (MAC). The authors showed that if the MAC to the legitimate receiver is “better”, in a precise sense, than the one to the adversary, then the legitimate users can reliably communicate on the order of \(\Theta(\sqrt{n})\) bits per \(n\) channel uses with arbitrarily LPD without using a secret key. The authors also quantified the pre-constant terms exactly. In \([23]\), the authors considered a BC communication model. However, note that the model in \([23]\) is significantly different from that in the present work. In \([23]\), the authors were interested in transmitting two messages, one common and another covert, over a BC where there are two receivers, one legitimate and the other the warden. The common message is to be communicated to both parties, while the covert message has LPD from the perspective of the warden. This models the scenario of embedding covert messages in an innocuous codebook and generalizes existing works on covert communications in which innocent behavior corresponds to lack of communication between the legitimate receivers. In our work, there are three receivers, two legitimate and the other denotes the warden. Both messages that are communicated can be considered as covert message from the perspective of the warden. See Fig. 1 for our model.

B. Main Contribution

Our main contribution is to establish the covert capacity region (the set of all achievable pre-constants of the throughputs which scale as \(\Theta(\sqrt{n})\)) of some two-user memoryless BCs with a warden when the no-input symbol is not redundant. While the (usual) capacity region of the two-user discrete memoryless BC is a long-standing open problem in network information theory, we show that under the covert communications constraint, the covert capacity region admits a particularly simple expression for a class of BCs we explicitly identify (see Condition 1). This region implies that time-division transmission is optimal for this class of two-user BCs. We emphasize the BC does not have to be degraded, less noisy or more capable \([10]\), Chapter 5]. Our main result is somewhat analogous to that of Lapidoth, Telatar and Urbanke \([24]\) who showed that time-division is optimal for wide-band broadcast communication over Gaussian, Poisson, “very noisy” channels, and average-power limited fading channels. However, the analysis in \([24]\) is restricted to stochastically degraded BCs.

The first analytical tool that is used to prove our main theorem is a converse bound derived by El Gamal \([25]\) initially designed for more capable BCs. However, it turns out to be a useful starting point for our setting. We manipulate this bound into a form that is reminiscent of an outer bound for degraded BCs. Subsequently, we express the \(\lambda\)-sum throughput of this outer bound, for appropriately chosen \(\lambda \geq 1\), in terms of upper concave envelopes \([26]\). This circumvents the need to identify the optimal auxiliary random variable for specific BCs. The identification of optimal auxiliaries is typically possible only

Fig. 1. Illustration of two-user BC with a warden. The key which is accessible to \(f, \varphi_1, \) and \(\varphi_2\) is not shown here.
mission is optimal for some BCs with covertness constraints.

C. Intuition Behind The Main Result

We now provide intuition as to why time-division transmission is optimal for some BCs with covertness constraints. Since symmetric channels satisfy our condition (point 3 in the discussion following Condition 1), we use the binary symmetric BC [10] Example 5.3, degraded in favor of $Y_1$ (say), as an example to illustrate this. Let $Y_j = X \oplus \Psi_j$ where $\Psi_j \sim \text{Bern}(p_j)$ for $j = 1, 2$, i.e., the channel from $X$ to $Y_j$ is a binary symmetric channel (BSC). It is well known that in the absence of the covertness constraint, superposition coding [10] Sec. 5.3 [30] is optimal, and time-division transmission is strictly suboptimal (unless the transmitter communicates with only a single receiver). The coding scheme is illustrated in Fig. 2 where $u_1^1(w_1), u_2^2(w_2) \in \mathbb{F}_2^n$ are generated independently, $u_2^2(w_2)$ denotes the cloud center and $x_n(w_1, w_2) = u_1^1(w_1) \oplus u_2^2(w_2)$ denotes the satellite codeword. Because in covert communications, $x_n(w_1, w_2)$ is required to have low (Hamming) weight [8], [9], either both $u_1^1(w_1)$ and $u_2^2(w_2)$ have low weight or both have high weight. Let us assume it is the former without loss of generality. Then, the set of locations of the 1’s in the satellite codeword is likely to be a disjoint union of the sets of 1’s in $u_1^1(w_1)$ and $u_2^2(w_2)$. Thus the relative weight of $x_n(w_1, w_2)$, denoted as $\alpha_n$, can be decomposed into the relative weights of $u_1^1(w_1)$ and $u_2^2(w_2)$, denoted as $\rho \alpha_n$ and $(1 - \rho)\alpha_n$ respectively for some $\rho \in [0, 1]$. Thus, the superposition coding inner bound for the degraded BC [10] Theorem 5.2 [30] with $X = U_1 \oplus U_2$ reads

\begin{align*}
R_1 & \leq I(X; Y_1|U_2) = I(U_1; U_1 \oplus \Psi_1) \approx \rho \alpha_n L_1^*, \\
R_2 & \leq I(U_2; Y_2) = I(U_2; U_2 \oplus \tilde{\Psi}_2) \approx (1 - \rho)\alpha_n L_2^*,
\end{align*}

where $L_1^*$ and $L_2^*$ are the “covert capacities” [8], [9] of the constituent point-to-point binary symmetric channels (see Theorem 1) and $\tilde{\Psi}_2 := U_1 \oplus \Psi_2$ is a Bernoulli random variable with parameter $\rho \alpha_n \cdot p_2 \approx p_2$ due to the low weight constraint on $U_1$ (its relative weight is $\rho \alpha_n$). Thus, (1) and (2) suggest that time-division is optimal, at least for symmetric BCs but we show the same is true for a larger class of BCs.

D. Paper Outline

This paper is structured as follows. In Section II we formulate the problem and define relevant quantities of interest precisely. In Section III we state our assumptions and main results. We also provide some qualitative interpretations. The main result, corollaries, and bounds on the required length of the secret key are proved in Sections IV, V, and VI respectively. We conclude our discussion in Section VII where we also suggest avenues for future research.

E. Notation

We adopt standard information-theoretic notation, following the text by El Gamal and Kim [10] and on occasion, the book by Csiszár and Körner [31]. We use $h_1(q) := -q \log q - (1 - q) \log(1 - q)$ to denote the binary entropy function. $D(P \| Q) := \sum_z P(z) \log \frac{P(z)}{Q(z)}$ and $\mathbb{V}(P, Q) := \frac{1}{2} \sum_z |P(z) - Q(z)|$ are the relative entropy and the total variation distance respectively. We use $I(X; Y)$ and $I(P, W)$ to denote the mutual information of $(X, Y) \sim P \times W$. Throughout $\log$ is taken to an arbitrary base. We also use $h(\cdot)$ to denote the differential entropy. The notation $a * b := a(1 - b) + b(1 - a)$ denotes the binary convolution operator. For two discrete distributions, $P$ and $Q$ defined on the same alphabet $\mathcal{X}$, we say that $P$ is absolutely continuous with respect to $Q$, denoted as $P \ll Q$ if for all $x \in \mathcal{X}$, $Q(x) = 0$ implies that $P(x) = 0$. Other notation will be introduced as needed in the sequel.
II. PROBLEM FORMULATION

A discrete memoryless two-user BC with a warden \((X, Y_1, Y_2, Z, P_{Y_1, Y_2, Z|X})\) consists of a channel input alphabet \(X\), three channel output alphabets \(Y_1, Y_2\), and \(Z\), and a transition matrix \(P_{Y_1, Y_2, Z|X}\). The output alphabets \(Y_1\) and \(Y_2\) correspond to the two legitimate receivers and \(Z\) corresponds to that of the warden. Without loss of generality, we let \(X = \{0, 1, \ldots, K\}\). We let \(0 \in X\) be the “no input” symbol that is sent when no communication takes place and define \(Q_x = P_{Z|X}(\cdot|x)\) for each \(x \in X\). The BC is used \(n\) times in a memoryless manner. If no communication takes place, the warden at receiver \(Z\) observes \(Z^n\), which is distributed according to \(Q_0^n\), the \(n\)-fold product distribution of \(Q_0\). If communication occurs, the warden observes \(Q^n\), the output distribution induced by the code. For convenience, in the sequel, we often denote the two marginal channels corresponding to the two legitimate receivers as \(W = P_{Y_1|X}\) and \(V = P_{Y_2|X}\) respectively.

The transmitter and the receivers are assumed to share a secret key \(S\) uniformly distributed over a set \(K\). We will, for the most part, assume that the key is sufficiently large, i.e., the set \(K\) is sufficiently large. However, we will bound the length of the key in Section III-E. The transmitter and the receiver aim to construct a code that is both reliable and covert. Let the messages to be sent be \(W_1\) and \(W_2\). These messages are assumed to be independent and also independent of \(S\). Also let their reconstructions at the receiver \(Y_j\) be \(\hat{W}_j\) for \(j = 1, 2\).

As usual, a code is said to be reliable if the probability of error \(Pr(\cup_{j=1}^2 \{W_j \neq \hat{W}_j\})\) vanishes as \(n \to \infty\). The code is covert if it is difficult for the warden to determine whether the transmitter is sending a message (hypothesis \(H_1\)) or not (hypothesis \(H_0\)). Let \(\pi_{1|0}\) 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 (one with no side information) satisfies \(\pi_{1|0} + \pi_{0|1} = 1\). The warden’s optimal hypothesis test satisfies

\[
\pi_{1|0} + \pi_{0|1} = 1 - \sqrt{\frac{D(Q^n_0)}{D(Q^n_2)}}
\]

where \(\sqrt{\cdot}\) follows Pinsker’s inequality; see [8, 12]. Hence, covertness is guaranteed if the relative entropy between the observed distribution \(Q^n_2\) and the product of no communication distribution \(Q^n_0\) is bounded by a small \(\delta > 0\).

Note that if \(\text{supp}(P_{Z|X}(\cdot|x)) \subseteq \text{supp}(Q_0)\) for some \(x \in X\), such \(x\) should not be transmitted, otherwise it is not possible for \(D(Q_n^Z || Q_n^n)\) to vanish [9]. Hence, by dropping all such input symbols as well as all output symbols not included in \(\text{supp}(Q_0)\), we assume throughout that \(\text{supp}(Q_0) = Z\). In addition, we assume that the no-input symbol \(0\) is not redundant i.e., \(P_{Z|X}(\cdot|0) \not\subseteq \text{conv}\{P_{Z|X}(\cdot|x'): x' \in X, x' \neq 0\}\) where \(\text{conv}\{\cdot\}\) denotes the convex hull. If the symbol 0 is redundant, there exists a sequence of codes for which \(D(Q_n^Z || Q_n^n) = 0\) for all \(n\) [9] so \(\pi_{1|0} + \pi_{0|1} = 1\) (i.e., the warden’s test is always blind) and transmitting at positive rates is possible; this is a regime we do not consider in this paper.

A code for the BC with a covertness constraint and the covert capacity region are defined formally as follows.

Definition 1. An \((n, M_{1n}, M_{2n}, K, \varepsilon, \delta)-code\) for the BC with a warden \((X, Y_1, Y_2, Z, P_{Y_1, Y_2, Z|X})\) and with a covertness constraint consists of

- Two message sets \(M_j := \{1, \ldots, M_{jn}\}\) for \(j = 1, 2\);
- Two independent messages uniformly distributed over their respective message sets, i.e., \(W_j \sim \text{Unif}(M_j)\) for \(j = 1, 2\);
- One secret key set \(K := \{1, \ldots, K\}\);
- One encoder \(f : M_{1n} \times M_{2n} \times K \to X^n\);
- Two decoders \(\varphi_j : Y_j^n \times K \to M_j\) for \(j = 1, 2\);

such that the following constraints hold:

\[
\Pr(\cup_{j=1}^2 \{W_j \neq \hat{W}_j\}) \leq \varepsilon, \quad D(Q_n^Z || Q_n^n) \leq \delta.
\]

For most of our discussion, we ignore the secret key set (i.e., we assume that the secret key is sufficiently long) for the sake of simplicity and refer to the family of codes above with secret key sets of arbitrary sizes as \((n, M_{1n}, M_{2n}, \varepsilon, \delta)\)-codes. We will revisit the effect of the key size in Section III-E.

Definition 2. We say that the pair \((L_1, L_2) \in \mathbb{R}^2_+\) is \((\varepsilon, \delta)\)-achievable for the BC with a warden and with a covertness constraint if there exists a sequence of \((n, M_{1n}, M_{2n}, \varepsilon, \delta)\)-codes such that

\[
\liminf_{n \to \infty} \frac{1}{\sqrt{n\delta}} \log M_{jn} \geq L_j, \quad j \in \{1, 2\},
\]

\[
\limsup_{n \to \infty} \varepsilon_n \leq \varepsilon.
\]

Define the \((\varepsilon, \delta)\)-covert capacity region \(L_{\varepsilon, \delta} \subset \mathbb{R}^2_+\) to be the closure of all \((\varepsilon, \delta)\)-achievable pairs of \((L_1, L_2)\). We are interested in the \(\delta\)-covert capacity region

\[
L_\delta := \bigcap_{\varepsilon \in (0,1)} L_{\varepsilon, \delta} = \lim_{\varepsilon \to 0} L_{\varepsilon, \delta}.
\]

Note that \(L_1\) and \(L_2\) in (7) are measured in bits (or information units) per square root channel use.

We will also need the notion of covert capacities for point-to-point discrete memoryless channels (DMCs) with a warden \((X, Y, Z, P_{Y,z|X})\) (8, 9). This scenario corresponds to the above definitions with \(Y_1 = Y\) and \(Y_2 = 0\). Recall that the chi-squared distance between two distributions \(Q_0\) and \(Q_1\) supported on the same alphabet \(Z\) is defined as

\[
\chi_2(Q_1 \parallel Q_0) := \sum_{z \in Z} \frac{(Q_1(z) - Q_0(z))^2}{Q_0(z)}.
\]

Theorem 1 (Bloch [8] and Wang, Wornell, Zheng [9]). Let \((X, Y, Z, P_{Y,z|X})\) be a DMC with a warden in which \(W := P_{Y|X}\) and \(Q_k := P_{Z|X}(\cdot|k)\) for \(k \in X\). We assume that it
satisfies \( W(\cdot|k) \ll W(\cdot|0) \) for all \( k \in \mathcal{X}\setminus\{0\} \), \( \text{supp}(Q_0) = \mathcal{Z} \) and \( 0 \in \mathcal{X} \) is not redundant. Then its covert capacity is

\[
L^*(P_{Y,Z|X}) := \max_P \sqrt{\frac{2(\sum_{k=1}^{K} p_k D(W(\cdot|k)\|W(\cdot|0)))^2}{\chi_2(\sum_{k=1}^{K} p_k Q_k\|Q_0)}} \tag{11}
\]

where the maximization extends over all length-\( K \) probability vectors \( \pmb{p} \) (i.e., \( \pmb{p} = [p_1, \ldots, p_K]^T \) where \( p_k \geq 0 \) and \( \sum_{k=1}^{K} p_k = 1 \)). If \( \mathcal{X} = \{0, 1\} \), i.e., \( P_{Y,Z|X} \) has a binary input, then the maximization over \( \pmb{p} \) in (11) is unnecessary and

\[
L^*(P_{Y,Z|X}) := \sqrt{\frac{2D(W(\cdot|1)\|W(\cdot|0))^2}{\chi_2(Q_1\|Q_0)}}. \tag{12}
\]

As previously mentioned, we assume that \( Q_1 \ll Q_0 \) in the binary input case (or more generally, \( \sum_{k=1}^{K} p_k^* Q_k \ll Q_0 \) where \( \pmb{p}^* = [p_1^*, \ldots, p_K^*]^T \) is any maximizer of (11)). Otherwise, covert communication is impossible \[8, Appendix G\]. With this assumption and the fact that \( 0 \in \mathcal{X} \) is not redundant, \( L^*(P_{Y,Z|X}) \) as defined in (11) and (12) is finite.

### III. Main Results

In this section, we present our main results. In Section III-A we state a condition on BCs, which we call Condition 1. In Section III-B we state our main result assuming the BC satisfies Condition 1 and the key is of sufficiently long length. We interpret our result in Section III-C by placing it in context via several remarks. In Section III-D we specialize our result to two degraded BCs and show that standard techniques apply for such models. In Section III-E we extend our main result to the case where the key size is also a parameter of interest.

#### A. A Condition on BCs

In the following, we consider the following condition on BCs that allows us to show that time-division is optimal.

**Condition 1.** Fix a BC with a warden \( P_{Y_1,Y_2,Z|X} \). Let the covert capacities of \( P_{Y_1,Z|X} \) and \( P_{Y_2,Z|X} \) be \( L_1^* := L^*(P_{Y_1,Z|X}) \) and \( L_2^* := L^*(P_{Y_2,Z|X}) \) respectively.

- If \( L_1^* \geq L_2^* \), we assume that

\[
\max_{P_X} \frac{I(X;Y_1)}{I(X;Y_2)} \leq \frac{L_1^*}{L_2^*}. \tag{13}
\]

- Otherwise if \( L_2^* \geq L_1^* \), we assume that

\[
\max_{P_X} \frac{I(X;Y_2)}{I(X;Y_1)} \leq \frac{L_2^*}{L_1^*}. \tag{14}
\]

A few remarks concerning Condition 1 are in order.

1) Condition 1 which is easy to check numerically as the optimizations over \( P_X \) are over compact sets, neither subsumes nor is subsumed by degradedness or any other ordering of \( W = P_{Y_1|X} \) and \( V = P_{Y_2|X} \). That is, we can show that there exists some degraded BCs that do not satisfy Condition 1 and there are also non-degraded BCs that satisfy Condition 1.

2) Condition 1 is significantly simplified in the binary-input case. Let \( W = P_{Y_1|X}, V = P_{Y_2|X}, \) and \( W_\gamma(y) := W(y|x), V_\gamma(x) := V(x|y) \) be the bernoulli distribution with probability of \( 1 \) being \( \gamma \in [0, 1] \), i.e.,

\[
P_\gamma(x) := \begin{cases} 1 - \gamma & x = 0 \\ \gamma & x = 1 \end{cases}. \tag{15}
\]

Similarly define \( V_\gamma \). Then, one can use (12) and Lemma 1 (to follow) to show that (13) is equivalent to

\[
\frac{D(W_1\|W_0)}{D(V_1\|V_0)} \leq \min_{\gamma \in [0, 1]} \frac{D(W_\gamma\|W_0)}{D(V_\gamma\|V_0)}. \tag{16}
\]

Thus the verification of Condition 1 for binary-input BCs reduces to a line search over \( [0, 1] \).

3) To illustrate this condition, we consider the scenario in which \( W = BSC(p) \) where \( p \in \{0.01, 0.20\} \) and \( V \) is a (generally) asymmetric binary-input, binary-output channel such that the transition matrix reads

\[
\sum_{x \in \mathcal{X}} P_\gamma(x)W(y|x), y \in \mathcal{Y} \text{ where } P_\gamma \text{ is the Bernoulli distribution with probability of } 1 \text{ being } \gamma \in [0, 1], \text{i.e.,}
\]

\[
P_\gamma(x) := \begin{cases} 1 - \gamma & x = 0 \\ \gamma & x = 1 \end{cases}. \tag{15}
\]

Similarly define \( V_\gamma \). Then, one can use (12) and Lemma 1 (to follow) to show that (13) is equivalent to

\[
\frac{D(W_1\|W_0)}{D(V_1\|V_0)} \leq \min_{\gamma \in [0, 1]} \frac{D(W_\gamma\|W_0)}{D(V_\gamma\|V_0)}. \tag{16}
\]

Thus the verification of Condition 1 for binary-input BCs reduces to a line search over \( [0, 1] \).

#### B. Time-Division is Optimal for Some BCs

Our main result is a complete characterization of the \( \delta \)-covert capacity region for all BCs satisfying Condition 1 and certain absolute continuity conditions.

**Theorem 2.** Assume that a BC with a warden \( P_{Y_1,Y_2,Z|X} \) is such that Condition 1 is satisfied and the constituent DMCs \( W := P_{Y_1|X} \) and \( V := P_{Y_2|X} \) satisfy \( W(\cdot|k) \ll W(\cdot|0) \) and \( V(\cdot|k) \ll V(\cdot|0) \) for all \( k \in \mathcal{X}\setminus\{0\} \). Also assume that the

\[2\]We use the notation \( W = BSC(q) \) if it is a binary-input, binary-output channel in which \( Y = X \oplus \Psi \) where \( \Psi \sim \text{Bern}(q) \).
length of the secret key is sufficiently large. Then, for all $\delta > 0$, the $\delta$-covert capacity region is
\[
\mathcal{L}_\delta = \left\{ (L_1, L_2) \in \mathbb{R}_+^2 : \frac{L_1}{L_1^1} + \frac{L_2}{L_2^2} \leq 1 \right\}.
\] (18)

Theorem 2 is proved in Section IV.

C. Remarks on the Main Theorem

A few remarks are in order.

1) First, note that (18) implies that under the covert communication constraints, time-division transmission is optimal for all BCs satisfying Condition 1. The achievability part simply involves two optimal covert communication codes, one for each DMC with a warden. The first code, designed for $P_{Y_1, Z_1|X}$, is employed over $[\rho n]$ channel uses where $\rho \in [0, 1]$. The second code, designed for $P_{Y_2, Z_2|X}$, is employed over the remaining $n - [\rho n]$ channel uses. However, because the normalization of $\log M_{jn}$ is $\sqrt{n}\delta$, we need a slightly more subtle time-division argument. To do so, fix $\delta' < \delta$, then from Theorem 1 we know that there exists a sequence of codes that allows transmission of
\[
\log M_{1n} \cong \sqrt{\rho n\delta L_1^*}
\] (19)
bits for user 1 over $[\rho n]$ channel uses and with covertness constraint (upper bound of the divergence in (6)) $\delta'$ and
\[
\log M_{2n} \cong \sqrt{(1-\rho)n(\delta-\delta')}L_2^*
\] (20)
bits for user 2 over $n - [\rho n]$ channel uses and with covertness constraint $\delta - \delta'$. Choosing $\delta'' = \rho\delta$ and combining these two codes, achieves the covertness constraint $\delta$ and the rate point $(\rho L_1^*, (1-\rho)L_2^*)$ which is on the boundary of $\mathcal{L}_\delta$. By varying $\rho \in [0, 1]$, we achieve the whole boundary and hence the entire region in (18).

Note that time-division is strictly suboptimal for the vast majority of BCs in the absence of the covert communication constraint. Indeed, one needs to perform superposition coding [30] to achieve all points in the capacity region for degraded, less noisy and more capable BCs. Thus, the covert communication constraint significantly simplifies the optimal coding scheme for BCs satisfying Condition 1.

2) The converse of Theorem 2 thus constitutes the main contribution of this paper. To obtain an explicit outer bound for the capacity region for BCs that satisfy some ordering—such as degraded, less noisy or more capable BCs [10, Chapter 5]—one often has to resort to the identification of the optimal auxiliary random variable-channel input pair $(U, X)$ in the capacity region of these classes of BCs. However, this is only possible for specific BCs; see Corollaries 1 and 2 to follow. For general (or even arbitrary degraded) BCs, this is, in general, not possible. Our workaround involves first starting with an outer bound of the capacity region for general memoryless BCs by El Gamal [25]. We combine the inequalities in the outer bounds and use this to upper bound a linear combination of the two throughputs in terms of the concave envelope of a linear combination of mutual information terms [26]. This allows us to circumvent the need to explicitly characterize $(U, X)$ since $U$ is no longer present in this concave envelope characterization. We then exploit Condition 1 and some approximations to obtain the desired outer bound to (18). The appeal of this approach is not only that we do not need to find the optimal $(U, X)$.

3) Let us say that the absolute continuity condition holds for $W$ if $W(\cdot|k) \ll W(\cdot|0)$ for all $k \in X_0 \setminus \{0\}$. Then Bloch [8, Appendix G] showed that if $W$ does not satisfy the absolute continuity condition, $\Theta(\sqrt{n}\log n)$ bits per $n$ channel uses can be covertly transmitted. In our setting, the same is true for the constituent channels; if $W$ satisfies the absolute continuity condition but $V$ does not, $\Theta(\sqrt{n})$ bits per $n$ channel uses can be transmitted to $Y_1$, while $\Theta(\sqrt{n}\log n)$ bits per $n$ channel uses can be transmitted to $Y_2$.

4) Finally, suppose that the covertness condition in (6) is replaced by a total variation constraint of the form
\[
\forall (\hat{Q}_{Z^n}, Q^n_0) \leq \delta, \quad \text{for some } \delta \in (0, 1).
\] (21)

Then, by using the techniques to prove [33, Theorem 2], one can easily see that the covert capacity region is the same as (18) in Theorem 2 except that in the formulation (7) is replaced by
\[
\lim_{n \to \infty} \frac{1}{\sqrt{n}\delta} \log M_{jn} \geq L_j, \quad j \in \{1, 2\},
\] (22)
where $\Gamma_k := \sqrt{2}\text{Q}^{-1}(1-\frac{1}{2})$ and $\text{Q}(.)$ is the complementary cumulative distribution function of a standard Gaussian. Hence, only the normalization (or scaling) is different. Since the justification of this is completely analogous to that of (the first-order result of) [33, Theorem 2], we omit the proof for the sake of brevity.

D. Two Degraded BCs

We now consider two specific classes of degraded BCs and show that modifications of standard techniques are applicable in establishing the outer bound to $\mathcal{L}_\delta$.

Corollary 1. Suppose that $W = \text{BSC}(p_1)$ and $V = \text{BSC}(p_2)$ with $0 \leq p_1 \leq p_2 \leq 1/2$ (without loss of generality) [10, Example 5.3]. Then $\mathcal{L}_\delta$ is as in (18) with
\[
L_j^* = (\log e)(1-2p_j)(\log \frac{1-p_j}{p_j}) \sqrt{\frac{2}{X_2(Q_1||Q_0)}}, \quad j \in \{1, 2\}.
\] (23)

An converse proof of this result follows from Mrs. Gerber’s lemma [27] and is presented in Section V.A. This result generalizes [9, Example 3].

We now consider the scenario in which the BC $P_{Y_1, Y_2, Z|X}$ consists of three additive white Gaussian noise (AWGN) channels [10, Sec. 5.5], i.e.,
\[
Y_j = X + \Psi_j, \quad j \in \{1, 2\}
\] (24)
\[
Z = X + \Psi_3,
\] (25)
where $\Psi_1, \Psi_2,$ and $\Psi_3$ are independent zero-mean Gaussian noises with variances $N_1, N_2,$ and $\sigma^2$ respectively. There is no peak, average, or long-term power constraint on the codewords [9, Sec. V]. Let the “no communication” input
In this section, we derive fundamental limits on the length of the secret key. The additional arguments to complete the proof of Theorem 3 are provided in Section VI.

The converse proof follows from the entropy power inequality [28], [29] and is presented in Section V-B. This result generalizes [8, Theorem 6] and [9, Theorem 5]. It is also analogous to [24, Prop. 1] in which it was shown that time-division is optimal for Gaussian BCs in the low-power limit.

We note that Corollaries 1 and 2 apply to an arbitrary but finite number of successively degraded legitimate receivers [10 Sec. 5.7], say \( N \). This means that \( X \rightarrow Y_1 \rightarrow Y_2 \rightarrow \ldots \rightarrow Y_N \). The corresponding \( \delta \)-covert capacity region is

\[
L_\delta = \left\{ (L_1, \ldots, L_N) \in \mathbb{R}^N_+ : \sum_{j=1}^N \frac{L_j}{\bar{L}_j} \leq 1 \right\},
\]

where \( \bar{L}_j \) is given as in (23) or (29). We omit the proofs as they are straightforward generalizations of the corollaries.

E. On the Length of the Secret Key

In the preceding derivations, we have assumed that the secret key that the transmitter and legitimate receivers share is arbitrarily long. In other words, the set \( K \) is sufficiently large. In this section, we derive fundamental limits on the length of the key so that covert communication remains successful.

To formalize this, we will need to augment Definition 2. We say that \((L_1, L_2, L_{key})\) is \( \delta \)-achievable or simply achievable if in addition to (7) and (8) (with \( \varepsilon = 0 \)),

\[
\lim_{n \to \infty} \frac{1}{\sqrt{n} \delta} \log K_n \leq L_{key}.
\]

Finally, we set \( L_n := L^* (P_{Z_n|X}) \). The generalization of Theorem 2 is as follows.

**Theorem 3.** Under the conditions of Theorem 2 the tuple \((L_1, L_2, L_{key})\) is achievable if and only if

\[
\frac{L_1}{L_1} + \frac{L_2}{L_2} \leq 1
\]

and

\[
L_{key} \geq \left( \frac{L_1}{L_1} + \frac{L_2}{L_2} \right) L^*_n - L_1 - L_2.
\]

Note that if the throughputs of the code (log \( M_{1n} \), log \( M_{2n} \)) are such that we operate on the boundary of the covert capacity region (i.e., that (29) holds with equality), the optimum (minimum) key length

\[
L^*_{key} = L^*_n - L_1 - L_2.
\]

Also note that if \( L_{key} \) is sufficiently large, then (30) is satisfied so based on (29), Theorem 3 reverts to Theorem 2. The proof of this enhanced theorem follows largely along the same lines as that for Theorem 2. However, we need to carefully bound the length of the secret key. The additional arguments to complete the proof of Theorem 3 are provided in Section VI.

IV. PROOF OF THEOREM 2

A. Preliminaries for the Proof of Theorem 2

Before we commence, we recap some basic notions in convex analysis. The (upper) concave envelope of \( f : \mathcal{D} \to \mathbb{R} \), denoted as \( \mathcal{C}(f)(x) := \inf \{ g(x) : g \geq f \text{ over } \mathcal{D}, g \text{ is concave} \} \), is the smallest concave function lying above \( f \). If \( \mathcal{D} \) is a subset of \( \mathbb{R}^d \), then by Carathéodory’s theorem [26],

\[
g = \mathcal{C}[f] \iff g(x) = \sup_{\{x_i, p_i\} : \sum_{i=1}^{d+1} p_i = 1} \sum_{i=1}^{d+1} p_i f(x_i).
\]

where \( x_i \in \mathcal{D} \) and \( \{p_i\}_{i=1}^{d+1} \) is a probability distribution such that \( \sum_{i=1}^{d+1} p_i x_i = x \). Here, we record a basic fact:

\[
f_2 \geq f_1 \text{ on } \mathcal{D} \implies g_2 \geq g_1 \text{ on } \mathcal{D}
\]

where \( g_j = \mathcal{C}[f_j], j = 1, 2. \) This can be shown by means of the representation of the concave envelope in (32). Indeed,

\[
g_2(x) \geq \sum_{i=1}^{d+1} p_i f_2(x_i) \geq \sum_{i=1}^{d+1} p_i f_1(x_i)
\]

where the first inequality holds for any \( \{x_i, p_i\}_{i=1}^{d+1} \) such that \( \sum_{i=1}^{d+1} p_i x_i = x \) and the second inequality because \( f_2 \geq f_1 \) on \( \mathcal{D} \). Since the inequality holds for all \( \{x_i, p_i\}_{i=1}^{d+1} \) such that \( \sum_{i=1}^{d+1} p_i x_i = x \), we can take the supremum of the right-hand-side of (34) over all such \( \{x_i, p_i\}_{i=1}^{d+1} \) to conclude that \( g_2 \geq g_1 \) on \( \mathcal{D} \).

B. Converse Proof of Theorem 2 Binary-Input BCs

We first prove the converse to Theorem 2 for the case when \( \mathcal{X} = \{0, 1\} \). This is done for the sake of clarity and simplicity. We subsequently show how to extend the analysis to the multiple symbol case (i.e., |\( \mathcal{X} \)| > 2) in Section IV-C.

Fix a sequence of \( (n, M_{1n}, M_{2n}, \varepsilon_n, \delta) \)-codes for the BC with a warden \( P_{Y_1, Y_2, Z_n|X} \) satisfying the \( \alpha \)-reliability constraint in (3) and (4) and the covertness constraint in (6). In the proof, we use the following result by Bloch [8, Lemma 1, Remark 1]:

**Lemma 1.** Let \( Q_{\gamma}(z) := \sum_{x \in \mathcal{X}} P_{Z|X}(z|x) P_{\gamma}(x) \) for \( \gamma \in [0, 1] \) where \( P_{\gamma} \) is defined in (15). Then, it follows that

\[
I(P_{\gamma}, P_{Z_n|X}) = \gamma D(Q_1||Q_0) - D(Q_\gamma||Q_0)
\]

Furthermore, for any sequence \( \gamma_n \) such that \( \gamma_n \to 0 \) as \( n \to \infty \), for all \( n \) sufficiently large,

\[
\frac{\gamma_n^2}{2} \chi^2(Q_1||Q_0) (1 - \gamma_n) \leq D(Q_\gamma||Q_0)
\]

\[
\leq \frac{\gamma_n^2}{2} \chi^2(Q_1||Q_0) (1 + \gamma_n^2).
\]

**1) Covertness Constraint:** We first discuss the covertness constraint in (6). Let \( Z_n^* \) (resp. \( X_{n+}^* \)) have distribution \( Q_{Z_n^*} \) (resp. \( P_{X_{n+}^*} \)) and let \( Q_{Z_n} \) (resp. \( P_{X_n} \)) be the marginal of \( Q_{Z_n} \) (resp. \( P_{X_n} \)) on the \( i \)-th element. Additionally, let \( Q_n \) be the average output distribution on \( Z \), i.e., \( Q_n := \frac{1}{n} \sum_{i=1}^{\infty} Q_i \). Similarly, we define the average input distribution \( P_n \) on \( X \) as \( P_n := \frac{1}{n} \sum_{i=1}^{\infty} P_{X_n} \). Then mimicking the steps in the proof of [9, Theorem 1], we have

\[
D(Q_{Z_n}||Q_0^{n+}) \geq n D(Q_n||Q_0).
\]
Thus, by the coverage constraint in (39) and (40), we have

\[ D(\hat{Q}_n||Q_0) \leq \frac{\delta}{n}. \]  
(38)

Since \( \lim_{n \to 0} D(\hat{Q}_n||Q_0) = 0 \) and symbol \( 0 \in \mathcal{X} \) is not redundant, it follows that

\[ \hat{P}_n = P_{\alpha_n} \text{ for some } \alpha_n \to 0. \]  
(39)

Because \( \hat{Q}_n(z) = Q_{\alpha_n}(z) \) for all \( z \in Z \), by Lemma 2,

\[ \frac{\alpha_n^2}{2} \chi_2(Q_1||Q_0)(1 - \sqrt{\alpha_n}) \leq D(\hat{Q}_n||Q_0) \]
\[ \leq \frac{\alpha_n^2}{2} \chi_2(Q_1||Q_0)(1 + \sqrt{\alpha_n}). \]  
(40)

From (38) and (40), we conclude that \( \alpha_n \) has to satisfy the weight constraint

\[ \alpha_n^2 (1 - \sqrt{\alpha_n}) \leq \frac{2\delta}{\chi_2(Q_1||Q_0)n} \Rightarrow \alpha_n^2. \]  
(41)

From this relation, we see that \( \alpha_n \leq \hat{\alpha}_n(1 + o(1)) = \Theta(\frac{1}{n}). \)

2) Upper Bound on Linear Combination of Code Sizes: We now proceed to the reliability and covertness constraints. Without loss of generality, we assume \( L^*_1 \geq L^*_2 \). We start with a lemma that is a direct consequence of the converse proof for more capable BCs by El Gamal [25]. This lemma is stated in a slightly different manner in [10, Theorem 8.5].

Lemma 2. Every \((n, M_{1n}, M_{2n}, \varepsilon_n)\)-code for any BC satisfies

\[ (\log M_{1n})(1 - \varepsilon_n) - 1 \leq \sum_{i=1}^n I(U_{1i}; Y_{1i}), \]  
(42)

\[ (\log M_{2n})(1 - \varepsilon_n) - 1 \leq \sum_{i=1}^n I(U_{2i}; Y_{2i}), \]  
(43)

\[ (\log M_{1n} + \log M_{2n})(1 - \varepsilon_n) - 2 \leq \sum_{i=1}^n [I(X_i; Y_{1i}|U_{2i}) + I(U_{2i}; Y_{2i})], \]  
(44)

\[ (\log M_{1n} + \log M_{2n})(1 - \varepsilon_n) - 2 \leq \sum_{i=1}^n [I(U_{1i}; X_i) + I(X_i; Y_{2i}|U_{1i})], \]  
(45)

where \( U_{1i} := (W_1, Y_{1i}^{i-1}, Y_{2i}^{i+1}) \) and \( U_{2i} := (W_2, Y_{1i}^{i-1}, Y_{2i}^{i+1}) \) satisfies \( (U_{1i}, U_{2i}) \rightarrow X_i \rightarrow (Y_{1i}, Y_{2i}). \) In addition, if a secret key \( S \) is available to the encoders and decoder, then the auxiliary random variables \( U_{1i} \) and \( U_{2i} \) also include \( S \).

For completeness, the proof of Lemma 2, with the effect of the secret key, is provided in Appendix A. Note that no assumption (e.g., degradedness, less noisy or more capable conditions) is made on the BC in Lemma 2.

\[ \text{Indeed, if } 0 \text{ were redundant (e.g., } P_{2i}|X(\cdot|0) = P_{2i}(\cdot|1) \text{ for binary } \mathcal{X}, \text{ there exists an input distribution } \hat{P} \text{ such that } \hat{P}(0) = 0 \text{ and its corresponding output distribution } \hat{Q} = \sum_x \hat{P}(x)P_{2i}|X(\cdot|x) = Q_0 \text{ that } \text{is trivially satisfied. However, } \hat{P} \text{ (taking the role of } P_n) \text{ clearly does not satisfy } 39. \]

Now fix a constant \( \lambda \geq 1 \). By adding \((\lambda - 1)\) copies of (43) to one copy of (44) and writing \( U_i := U_{2i} \) for all \( i \in \{1, \ldots, n\} \), we obtain

\[ (\log M_{1n} + \lambda \log M_{2n})(1 - \varepsilon_n) - (1 + \lambda) \]
\[ \leq \sum_{i=1}^n [I(X_i; Y_{1i}|U_i) + \lambda I(U_i; Y_{2i})]. \]  
(46)

By introducing the usual time-sharing random variable \( J \in \text{Unif}\{1, \ldots, n\} \), which is independent of all the other random variables \( (W_1, W_{2n}, X^n, Y_{1n}^n, Y_{2n}^n) \) and defining \( U := (J, U_J) \), \( X := X_J \), \( Y_1 := Y_{1J} \) and \( Y_2 := Y_{2J} \), we obtain

\[ (\log M_{1n} + \lambda \log M_{2n})(1 - \varepsilon_n) - (1 + \lambda) \]
\[ \leq n I(X; Y_1|U) + n \lambda I(U; Y_2). \]  
(47)

Note that the above identification of \((U, X)\) satisfies \( U \rightarrow X \rightarrow (Y_1, Y_2). \) Furthermore, since \( X \) is equal to \( X_i \) for each \( i \in \{1, \ldots, n\} \) with equal probability \( 1/n \) (because of \( X = X_J \)), it follows \( P_X = P_n = P_{\alpha_n} \) for some \( \alpha_n \) satisfying (41).

Hence, (47) can be bounded by taking a maximization over all such distributions, i.e.,

\[ \frac{1}{n} [(\log M_{1n} + \lambda \log M_{2n})(1 - \varepsilon_n) - (1 + \lambda)] \]
\[ \leq \max_{P_X} \max_{P_{U|X}} I(X; Y_1|U) + \lambda I(U; Y_2), \]  
(48)

where the maximization over \( P_X \) is over distributions \( P_{\alpha_n} \) where \( \alpha_n \) satisfies (41).

Now by using the concave envelope representation for the capacity region of degraded BCs [26], we may write (47) as follows:

\[ \frac{1}{n} [(\log M_{1n} + \lambda \log M_{2n})(1 - \varepsilon_n) - (1 + \lambda)] \]
\[ \leq \max_{P_X} \max_{P_{U|X}} I(X; Y_1|U) + \lambda I(U; Y_2) \]
\[ = \max_{P_X} I(X; Y_1|U) + \lambda [I(X; Y_2) - I(X; Y_2|U)], \]  
(49)

\[ \max_{P_X} \lambda I(X; Y_2) + \max_{P_{U|X}} I(X; Y_1|U) - \lambda I(X; Y_2|U), \]  
(50)

\[ \max_{P_X} \lambda I(X; Y_2) + \epsilon_{P_X} [I(X; Y_1) - \lambda I(X; Y_2)]. \]  
(51)

Here, (50) follows from the Markov chain \( U \rightarrow X \rightarrow (Y_1, Y_2) \) and (52) follows from the definition of the concave envelope. Namely, for fixed \( P_X, \)

\[ \max_{P_{U|X}} I(X; Y_1|U) - \lambda I(X; Y_2|U), \]  
(53)

\[ = \max_{P_{U|X}} \sum_{u} P_U(u)[I(X; Y_1|U = u) - \lambda I(X; Y_2|U = u)], \]  
(54)

\[ = \max_{P_{U|X}} \sum_{u} P_U(u)[I(P_X; U|\cdot(u), W) - \lambda I(P_X; U|\cdot(u), V)], \]  
(55)

where (54) follows from the Markov chain \( U \rightarrow X \rightarrow (Y_1, Y_2) \) and (55) follows from the fact that \( P_{X|U}(\cdot|u) = P_X \) and (52). Note that we employ the subscript \( P_X \) on \( \epsilon \) to emphasize that the concave envelope operation is taken with respect to the distribution \( P_X \) and it is thus a function of \( P_X \).
3) Approximating the Maximization over Low-Weight Inputs: Due to the above considerations, it now suffices to simplify (52). To obtain the outer bound to (13), we set \( \lambda := \frac{L_1^*}{L_2^*} \) in the following. Note that since it is assumed that \( L_1^* \geq L_2^*, \lambda \geq 1 \) as required to obtain (46). If instead \( L_2^* > L_1^* \), then we set \( \lambda := \frac{L_2^*}{L_1^*} \) and add \((\lambda - 1)\) copies of (42) to one copy of (45) to obtain the analogue of (46) with \( U_i := U_{1i} \). We also use (44) instead of (13) in Condition I. Finally, in the rest of the proof, we replace the index 1 by 2 and \( W \) by \( V \) and vice versa. The following arguments go through verbatim with these minor amendments.

By expressing the mutual information quantities in (52) as \( I(P_X, W) \) and \( I(P_X, V) \) (for \( I(X; Y_1) \) and \( I(X; Y_2) \) respectively), we can write (52) as follows:

\[
\max_{P_X} \lambda I(P_X, V) = \mathcal{E}_{P_X}[I(P_X, V) - \lambda I(P_X, V)].
\]

By applying (13) of Condition I, we see that \( I(P_X, W) - \lambda I(P_X, V) \leq 0 \) for all \( P_X \). Hence, by using the monotonicity property in (33) and the fact that the concave envelope of the 0 function is 0, we obtain that (56) is upper bounded by

\[
\max_{P_X} \lambda I(P_X, V) \geq \frac{\alpha_n}{n} \cdot D(\cdot | 1) || V(\cdot) || (57)
\]

Now we parametrize \( P_X = P_{\alpha_n} = \tilde{P}_n \) as the vector \([1 - \alpha_n, \alpha_n]^T\) where \( 0 \leq \alpha_n \leq (1 + o(1)) \) [cf. (41)]. Appealing to Lemma I, we see that the value of the optimization problem in (57) is given by

\[
\lambda \tilde{\alpha}_n (1 + o(1)) \cdot D(V(\cdot | 1) || V(\cdot)).
\]

Now, combining (56) with the upper bound in (52), we have

\[
\frac{1}{n} \left[ \log M_1 n + \lambda \log M_2 n \right] (1 - \varepsilon_n) - (1 + \lambda) \leq \lambda \tilde{\alpha}_n (1 + o(1)) \cdot D(V(\cdot | 1) || V(\cdot)).
\]

4) Completing the Converse Proof Taking Limits: At this point, we invoke the definition of \( \tilde{\alpha}_n \) in (41) and normalize (59) by \( \sqrt{n} \tilde{\alpha}_n \) to obtain

\[
\lambda (1 + o(1)) \sqrt{\frac{2 \delta n}{\chi^2(Q_1 || Q_0)} \cdot D(V(\cdot | 1) || V(\cdot))}.
\]

Taking the lim sup in \( n \) on both sides, recalling that (i) \( \lambda = \frac{L_1^*}{L_2^*} \geq 1 \) and (ii) the definition of achievable \((L_1, L_2)\) pairs according to Definition 2 and the facts that \( \varepsilon_n \to 0 \) and \( n \delta \to \infty \), we obtain

\[
L_1 + L_2^* \leq \frac{2}{\chi^2(Q_1 || Q_0)} \cdot D(V(\cdot | 1) || V(\cdot)).
\]

Finally, by recalling the definition of the covert capacity of \( P_{Y_2, Z| X} \) according to (12) in Theorem 1, we see that the right-hand-side of (61) is exactly \( L_1^* \). Hence we obtain the desired outer bound corresponding to \( L_\delta \) in (18) in Theorem 2.

C. Converse Proof of Theorem 2: BCs with Multiple Inputs

We generalize the converse proof for binary-input in Section IV-B to the multiple symbol case, i.e., \( |X| > 2 \). We use the following lemma whose proof is in [8] Section VII-B.

Lemma 3. For \( \gamma \in [0, 1] \) and length-\( \gamma \) probability vector \( p = [p_1, \ldots, p_K]^T \), i.e., \( p_k \geq 0 \) and \( \sum_{k=1}^K p_k = 1 \), let \( Q_\gamma, p = (z|x)P_\gamma, p(x) \) where

\[
P_\gamma, p(x) := \begin{cases} 1 - \gamma_p & x = 0 \\ \gamma_p & x = k \in \mathcal{X} \setminus \{0\} \end{cases}.
\]

Then, it follows that

\[
I(P_\gamma, p, Z|X) = \left( \sum_{k=1}^K p_k Q_k \right) - D(Q_\gamma, p, Q_0).
\]

Furthermore, for any sequence \( \gamma_n \) such that \( \gamma_n \to 0 \) as \( n \to \infty \), for all \( n \) sufficiently large,

\[
\frac{\gamma_n^2}{2} \chi_2 \left( \sum_{k=1}^K p_k Q_k \right) \leq D(Q_{\gamma_n}, p, Q_0) \leq \frac{\gamma_n^2}{2} \chi_2 \left( \sum_{k=1}^K p_k Q_k \right) (1 + \sqrt{\gamma_n})
\]

1) Coverness Constraint: By a close inspection of the steps in Section IV-B and by applying Lemma 3, we conclude that the average input distribution \( P_n \) is given by \( P_{\alpha_n, p} \) for some length-\( \gamma \) probability vector \( p \) and \( \alpha_n \in [0, 1] \) satisfying (65). As such, we have that (52) also holds where now the maximization over \( P_X \) is over those distributions \( P_{\alpha_n, p} \) for some \( p \) and \( \alpha_n \) satisfying (65).

2) Upper Bound on Linear Combination of Code Sizes: Let us assume \( L_1^* \geq L_2^* \) (without loss of generality), then \( \lambda = \frac{L_1^*}{L_2^*} \geq 1 \). It can be easily checked that the steps in Section IV-B go through in the same manner for the multiple symbol case except that \( P_X = P_{\alpha_n, p} \) for some length-\( \gamma \) probability vector \( p \) and \( \alpha_n \in [0, 1] \) satisfying (65). As such, we have that (52) also holds where now the maximization over \( P_X \) is over those distributions \( P_{\alpha_n, p} \) for some \( p \) and \( \alpha_n \) satisfying (65).

3) Approximating the Maximization over Low-Weight Inputs: Now we approximate (56). We fix \( p \) throughout. By Condition 1, \( I(P_X, W) - \lambda I(P_X, V) \leq 0 \) and so only the first term \( \lambda I(P_X, V) \) remains in the maximization. Similarly to (56), the maximization in (57) can be approximated by

\[
\lambda \tilde{\alpha}_{p, n} (1 + o(1)) \sum_{k=1}^K p_k D(V(\cdot | k) || V(\cdot))
\]

as \( n \to \infty \).

Due to the same reasoning up to (59), we have

\[
\frac{1}{n} \left[ \log M_1 n + \lambda \log M_2 n \right] (1 - \varepsilon_n) - (1 + \lambda) \leq \lambda \tilde{\alpha}_{p, n} (1 + o(1)) \sum_{k=1}^K p_k D(V(\cdot | k) || V(\cdot))
\]

as \( n \to \infty \).
4) Completing the Converse Proof Taking Limits: By substituting \( \lambda = L_1^*/L_2 \) and \( \alpha_{p, n} \) defined in (65) into (67), taking limits as in Section V-B.4 and subsequently maximizing over all probability vectors \( \mathbf{p} \) on \( \mathcal{X} \setminus \{0\} \), we obtain

\[
L_1 + \frac{L_1^*}{L_2} \leq \frac{L_1^*}{L_2} \max_{\mathbf{p}} \left[ \frac{2(\sum_{k=1}^K p_k D(V(\cdot|k)||V(\cdot|0)))^2}{\chi^2(\sum_{k=1}^K p_k Q_k||Q_0)} \right] = L_1^*. \tag{68}
\]

This completes the proof for the multiple symbol case.

V. Proofs of Corollaries in Section III-D

A. Proof of Corollary 1

As the capacity region of the stochastically degraded BC is the same as that of the physically degraded BC with the same marginals [10, Sec. 5.4] even under the covertness constraint, we assume that the physically degradedness condition holds, i.e., \( Y_1 = X \oplus \Psi_1, Y_2 = Y_1 \oplus \Psi_2 \) where \( \Psi_2 \sim \text{Bern}(\tilde{p}_2) \) such that \( p_1 \ast \tilde{p}_2 = \tilde{p}_2 \).

We follow the exposition in [10, Sec. 5.4.2] closely. By the standard argument to establish the weak converse for degraded BCs [34], we know that every \( (n, M_{1n}, M_{2n}, \varepsilon_n) \)-code for a BC must satisfy

\[
(\log M_{1n})(1 - \varepsilon_n) - 1 \leq n I(X; Y_1|U) \tag{69}
\]

\[
(\log M_{2n})(1 - \varepsilon_n) - 1 \leq n I(U; Y_2), \tag{70}
\]

for some \( U \rightarrow X \rightarrow (Y_1, Y_2) \) in which \( X \) satisfies the weight constraint in (72), i.e., \( P_X = P_{\alpha_n} \) or \( \mathbb{E}[X] = \alpha_n \). We now evaluate the region in (69–70) explicitly in terms of the channel parameters and the upper bound on the weight of \( X \).

Now we have

\[
h_b(p_2) = H(Y_2|X) \geq h_b(\alpha_n * p_2), \tag{73}
\]

where the equality in (72) follows from the Markov chain \( U \rightarrow X \rightarrow (Y_1, Y_2) \) and the fact that the entropy of \( Y_2 = X \oplus \Psi_1 \oplus \Psi_2 \) when \( \mathbb{E}[X] = \alpha_n \) is \( h_b(\alpha_n * p_2) \). As a result, from (73), there exists a \( \tau_n \in [0, \alpha_n] \) such that

\[
H(Y_2|U) = h_b(\tau_n * p_2). \tag{74}
\]

Then, we have

\[
h_b(\tau_n * p_2) = h_b(\tau_n * p_2) \geq h_b(\tilde{p}_2, (H(Y_1|U)) * \tilde{p}_2) \tag{75}
\]

where (76) follows from the conditional version of Mrs. Gerber’s lemma [27]. By the monotonicity of \( h_b(\cdot) \) on \( [0, 1/2] \) and the fact that \( p_1 \ast \tilde{p}_2 = \tilde{p}_2 \), we have

\[
\tau_n * p_1 \ast \tilde{p}_2 \geq h_b^{-1}(H(Y_1|U)) * \tilde{p}_2. \tag{77}
\]

As a result, we have

\[
H(Y_1|U) \leq h_b(\tau_n * p_1). \tag{78}
\]

Now consider

\[
I(X; Y_1|U) = H(Y_1|U) - H(Y_1|U, X) \tag{79}
\]

\[
= H(Y_1|U) - h_b(p_1) \tag{80}
\]

\[
\leq h_b(\tau_n * p_1) - h_b(p_1). \tag{81}
\]

For the other rate bound in (70), we have

\[
I(U; Y_2) = H(Y_2) - H(Y_2|U) \tag{82}
\]

\[
\leq h_b(\alpha_n * p_2) - h_b(\tau_n * p_2). \tag{83}
\]

Now, note that for all \( q \in (0, 1/2), \frac{d}{dq} h_b(q) = (\log e) \log \frac{1 - q}{q} \).

By Taylor expanding \( h_b \) around \( q \), we obtain

\[
h_b(q * \xi) = h_b(q) + (1 - 2q)\xi \tag{84}
\]

\[
= h_b(q) + (\log e) \log \frac{1 - q}{q} \cdot (1 - 2q)\xi + O(\xi^2) \tag{85}
\]

as \( \xi \to 0 \). Now we can simplify the bounds on the right-hand-sides of (81) and (83). Write \( \alpha_n := a/\sqrt{n} \) and \( \tau_n := t/\sqrt{n} \) where \( t \in [0, a] \) and \( a := (1 + o(1))\sqrt{25/\chi^2(Q_1||Q_0)} \); see [41] for the relation between \( \alpha_n \) and an upper bound on a function of it given by \( \alpha_n \). By combining (69), (70), (81), (83) and (85), we obtain the outer bound

\[
\frac{1}{n} \left[ \log M_{1n} \right] \left(1 - \varepsilon_n - 1\right) \leq (\log e) \left(1 - 2p_1\right) \log \frac{1 - p_1}{p_1} \cdot \frac{t}{\sqrt{n}} + O \left(\frac{1}{n}\right) \tag{86}
\]

\[
\frac{1}{n} \left[ \log M_{2n} \right] \left(1 - \varepsilon_n - 1\right) \leq (\log e) \left(1 - 2p_2\right) \log \frac{1 - p_2}{p_2} \cdot \frac{a - t}{\sqrt{n}} + O \left(\frac{1}{n}\right). \tag{87}
\]

By dividing the first and second bounds by \( (\log e) \left(1 - 2p_1\right) \log \frac{1 - p_1}{p_1} \) and \( (\log e) \left(1 - 2p_2\right) \log \frac{1 - p_2}{p_2} \) respectively, adding them, multiplying the resultant expression by \( \sqrt{n}/\delta \), and taking limits, we immediately recover the outer bound to \( \mathcal{L}_\delta \) with \( L_1^* \) defined in (23).

B. Proof of Corollary 2

The proof follows similarly to that of Corollary 1 but we use the entropy power inequality [23, 29] in place of Mrs. Gerber’s lemma. As in the proof of Corollary 1 we assume physically degradedness, i.e., \( Y_1 = X + \Psi_1, Y_2 = Y_1 + \Psi_2 \) where \( \Psi_2 \) is independent zero-mean Gaussian noise with variance \( N_2 - N_1 \). Let the second moment of the input distribution be denoted as \( \alpha_n = \mathbb{E}[X^2] \); this plays the role of the weight of \( X \) in the discrete case. The bounds (69) and (70) clearly still hold so we only have to single-letterize the two mutual information terms \( I(X; Y_1|U) \) and \( I(U; Y_2) \). We follow the exposition in [10, Sec. 5.5.2] closely.

First, we have

\[
I(U; Y_2) = h(Y_2) - h(Y_2|U) \tag{88}
\]

\[
\leq \frac{1}{2} \log(2\pi e(\alpha_n + N_2)) - h(Y_2|U) \tag{89}
\]
where follows from the fact that a Gaussian maximizes differential entropy over all distributions with the same second moment. Now note that
\[
\frac{1}{2} \log(2\pi e N_2) = h(Y_2 | X)
\]
\[
= h(Y_2 | U, X) \leq h(Y_2 | U) \leq h(Y_2)
\]
\[
\leq \frac{1}{2} \log(2\pi e (\alpha_n + N_2)).
\]
As such there exists a \( \tau_n \in [0, \alpha_n] \) such that
\[
h(Y_2 | U) = \frac{1}{2} \log(2\pi e (\tau_n + N_2)).
\]
Then, by the conditional form of the entropy power inequality [28], [29],
\[
h(Y_2 | U) = h(Y_1 + \tilde{\Psi}_2 | U)
\]
\[
\geq \frac{1}{2} \log(2^{2h(Y_1 | U)} + 2^{2h(\tilde{\Psi}_2 | U)})
\]
\[
= \frac{1}{2} \log \left( 1 + \frac{\tau_n}{N_1} \right) \leq \frac{\tau_n \log e}{2N_1},
\]
where the last inequality follows from the fact that \( \log(1+t) \leq t \log e \) for \( t \geq -1 \). Similarly,
\[
I(U; Y_2) = h(Y_2 | U) - h(Y_2 | U, X)
\]
\[
\leq \frac{1}{2} \log \left( \frac{\alpha_n + N_2}{\tau_n + N_2} \right) \leq \frac{\alpha_n - \tau_n \log e}{2(\tau_n + N_2)}
\]
\[
\leq \frac{(\alpha_n - \tau_n) \log e}{2N_2}.
\]
Using the same calculations as those leading to [9] Eqn. (75)],, we conclude that the covert communication constraint in the Gaussian case translates to
\[
\frac{\alpha_n^2}{4\pi} + o(\alpha_n^2) \leq \frac{\delta}{n}.
\]
The proof is completed by uniting (100), (103), and (104) in a way that is analogous to the conclusion of the proof for the binary symmetric BC in Section VI-A.

VI. PROOF SKETCH OF THEOREM 3

Proof: The achievability in (30) follows by augmenting the standard time-division argument given in Remark 1 in Section III-C. More specifically, we let \( \rho_1 := L_1 / \tau_n^* \) and \( \rho_2 := L_2 / \tau_n^* \). Then \( \rho_1 + \rho_2 \leq 1 \). We use an optimal code for the first receiver for \( \rho_1 \) fraction of time, an optimal code for the second receiver for \( \rho_2 \) fraction of time, and stay idle for the remaining fraction of time. Then the required secret key rate is \( \rho_1 (L_1^* - L_1^*) + \rho_2 (L_2^* - L_2^*) \) as a direct consequence of Theorem 3, which is equal to the right-hand side of (30).

For the converse, for simplicity, we only consider the binary-input case; the general case follows from replacing \( Q \) in Theorem 3, and \( L \) everywhere. We show that there is a tradeoff between the message rates and the secret key rate (with normalizations \( \sqrt{n} \)); if the message rates are in the interior of (18), a strictly smaller secret rate is required. To elucidate this tradeoff, let us consider a rate tuple \( (L_1, L_2) \) such that \( L_1 / \tau_n^* + L_2 / \tau_n^* = \kappa \) for some \( \kappa \in [0, 1] \). Note that \( \kappa \) strictly less than one means that the message rates are in the interior of (18).

By applying similar steps as in [8] Eqs. (91)–(94), we obtain
\[
\log M_1 + \log M_2 + \log K_n = H(W_1, W_2, S)
\]
\[
\geq I(W_1, W_2, S; Z^n) \geq I(X^n; Z^n) \geq n I(X; Z) - \delta,
\]
where the final inequality in (106) follows from similar steps presented in [35] Sec. 5.2.3. Here \( X = X_I \) and \( Z = Z_J \) where \( J \) is the time-sharing random variable defined in Section IV-B2. Then,
\[
\frac{1}{\sqrt{n} \delta} (\log M_1 + \log M_2 + \log K_n)
\]
\[
\geq \frac{1}{\sqrt{n} \delta} (n I(X; Z) - \delta)
\]
\[
= \frac{n \alpha_n}{\sqrt{n}} \left( D(Q_1 || Q_0) - \frac{D(Q_0 || Q_0)}{\alpha_n} \right) - \frac{\delta}{n \alpha_n},
\]
where (108) follows from Lemma 1. Note that the lower bound (108) depends on the “weight” \( \alpha_n = \bar{P}_n(1) \). Hence, substituting the smallest “weight” \( \alpha_n = \bar{P}_n(1) \) that attains \( L_1 / \tau_n^* + L_2 / \tau_n^* = \kappa \) into (108) results in a lower bound on the required secret key rate to attain the message rate pair \( (L_1, L_2) \).

Now let us denote the smallest \( \alpha_n = \bar{P}_n(1) \) that attains \( L_1 / \tau_n^* + L_2 / \tau_n^* = \kappa \) by \( \nu \bar{\alpha}_n \). By a close inspection of the proof of Theorem 2, we can check that under the constraint \( \alpha_n = \nu \bar{\alpha}_n \), the message pairs are bounded by \( L_1 / \tau_n^* + L_2 / \tau_n^* \leq \nu \). Hence, to achieve \( L_1 / \tau_n^* + L_2 / \tau_n^* = \kappa \), it should follow that \( \nu \geq \kappa \). Indeed, we can see that \( L_1 / \tau_n^* + L_2 / \tau_n^* = \kappa \) is achievable with weight \( \kappa \bar{\alpha}_n \) from the direct part above, by communicating \( \kappa \) fraction of the time and staying idle for the remaining \( 1 - \kappa \) fraction.

Note that when \( \alpha_n = \kappa \bar{\alpha}_n \), it follows from (56) that
\[
\lim_{n \to \infty} \frac{D(Q || Q_0)}{\alpha_n} = 0 \quad \text{and} \quad \lim_{n \to \infty} \frac{\delta}{n \alpha_n} = 0.
\]
Then, by combining (108)–(109), we have
\[
\lim_{n \to \infty} \frac{1}{\sqrt{n} \delta} (\log M_1 + \log M_2 + \log K_n)
\]
\[
\geq \kappa \sqrt{\frac{2D(Q || Q_0)^2}{\chi_2(Q || Q_0)}} = \kappa L_2^*.
\]

4With a slight abuse of terminology, we will refer to \( L_1, L_2, \) and \( L_{\text{msg}} \) as rates even though they are not communication rates in the usual sense [10].

5We omit multiplying \( \bar{\alpha}_n \) by the factor \( 1 + o(1) \) as this is inconsequential in the (asymptotic) arguments that follow.
Since \( \kappa = L_1/L_1^* + L_2/L_2^* \), (110) immediately implies a matching converse to (30).

VII. Conclusion and Future Work

In this paper, we established the covert capacity region for two-user memoryless BCs that satisfy Condition [1]. Somewhat surprisingly, the most basic multi-user communication strategy—time-division transmission—turns out to be optimal for this class of BCs. Our proof strategy provides further evidence that the concave envelope characterization of bounds on capacity regions in network information theory [26] is convenient and useful.

There are at least three promising avenues for future work. 1) Because we are adopting the average probability of error formalism in (5), it is likely that the strong converse (in \( \varepsilon \)) does not hold as suggested by the argument in Appendix A of [33]. Verifying that the argument therein indeed extends to BCs would be a natural avenue for future work. In addition, proving that a strong converse holds under the maximum probability of error formalism would also be a fruitful endeavor.

2) What about the covert capacity region for BCs that do not satisfy Condition [1]? In this case time-division may not be optimal and a natural avenue for future work would be to construct schemes to beat the time-division inner bound for such BCs.

3) What is the covert capacity region for BCs (under appropriate conditions) in which there are more than two legitimate receivers? While Corollaries [1] and [2] hold for an arbitrary number of legitimate receivers (see (27), Theorem [2] and in particular Lemma [2] in which it hinges on, does not seem to generalize easily.

APPENDIX A

PROOF OF LEMMA [2]

Proof: We will only prove (42) and (44) as the other two bounds follow by swapping indices 1 with 2 and vice versa. In addition, in this proof, we include the effect of the secret key \( S \). By Fano’s inequality,

\[
H(W_j|Y_j^n, S) \leq \varepsilon_n \log M_{jn} + 1, \quad j \in \{1,2\}.
\]

Because \( W_1 \) and \( W_2 \) are uniform and independent of each other and of \( S \), we have

\[
(\log M_{jn}) (1-\varepsilon_n) - 1 \leq I(W_1; Y_1^n|S) \quad (112)
\]

\[
(\log M_{jn} + \log M_{jn}) (1-\varepsilon_n) - 2 \leq I(W_1; Y_1^n|W_2^n, S) + I(W_2; Y_2^n|S). \quad (113)
\]

We start by single letterizing (112) as follows:

\[
(\log M_{jn}) (1-\varepsilon_n) - 1 
\leq \sum_{i=1}^{n} I(W_1; Y_1^i|Y_1^{i-1}, S) \quad (114)
\]

\[
\leq \sum_{i=1}^{n} I(W_1, S, Y_1^{i-1}, Y_2^n, Y_2^{i-1}); Y_1^i) \quad (115)
\]

\[
= \sum_{i=1}^{n} I(U_1; Y_1^i), \quad (116)
\]

where (116) follows from the identification \( U_1 := (W_1, S, Y_1^i, Y_2^{i+1}). This proves (42).

Next we single letterize (113) using the steps at the top of the next page, where (121) follows from the Csiszár-sum-identity [10] and the identification \( U_{2i} := (W_2, S, Y_{1,i+1}, Y_{2,i}^{-1}) \) and (122) follows from the fact that \( X_i \) is a function of \( (W_1, W_2, S) \). This proves (44).

Acknowledgements

The authors thank Dr. Ligig Wang (ETIS—Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS) for helpful discussions in the initial phase of this work and for pointing us to [24]. The first author also thanks Dr. Lei Yu (NUS) for fruitful discussions on the relation of the present work to stealth [35]. The authors would like to thank the Associate Editor, Prof. Rainer Böhme, and the anonymous reviewers for their useful feedback to improve the quality of the manuscript.

REFERENCES

[1] V. Y. F. Tan and S.-H. Lee. Time-division is optimal for covert communication over some broadcast channels. In Proc. of the IEEE Inform. Th. Workshop, Guangzhou, China, Nov 2018.

[2] B. A. Bash, D. Goekel, and D. Towsley. Limits of reliable communication with low probability of detection on AWGN channels. IEEE J. Sel. Areas Commun., 31(9):1921–1930, Sep 2013.

[3] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha. Quantum-secure covert communication on bosonic channels. Nature Commun., 6:8626, Oct 2015.

[4] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha. Hiding information in noise: Fundamental limits of covert wireless communication. IEEE Communications Magazine, 53(12):26–31, Dec 2015.

[5] P. H. Che, M. Bakshi, and S. Jaggi. Reliable deniable communication: Hiding messages in noise. In Proc. IEEE Int. Symp. Inform. Theory, pages 2945–2949, Istanbul, Turkey, Jul 2013.

[6] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi. Reliable deniable communication with channel uncertainty. In Proc. of the IEEE Inform. Th. Workshop, pages 30–34, Hobart, Australia, Nov 2014.

[7] S. Lee, R. J. Baxley, M. A. Wientrnan, and B. Wallenhorst. Achieving undetectable communication. IEEE J. Sel. Topics Signal Process., 9(7):1195–1205, Oct 2015.

[8] M. R. Bloch. Covert communication over noisy channels: A resolvability perspective. IEEE Trans. Inf. Theory, 62(5):2334–2354, 2016.

[9] L. Wang, G. W. Wornell, and L. Zheng. Fundamental limits of communication with low probability of detection. IEEE Trans. Inf. Theory, 62(6):3493–3503, 2016.

[10] A. El Gamal and Y.-H. Kim. Network Information Theory. Cambridge University Press, Cambridge, U.K., 2012.

[11] R. Anderson. Stretching the limits of steganography. In Proc. 1st Information Hiding Workshop, pages 39–48, volume 1174 of Springer LNCS, 1996.

[12] A. D. Ker, P. Levny, J. Kodovský, and J. Fridrich. The square root law of steganographic capacity. Proceedings of the 10th ACM Multimedia & Security Workshop, pages 107–116, 2008.

[13] V. Kortik, G. Morales-Luna, and M. H. Lee. On the existence of perfect stegosystems. In Proc. 4th Int. Workshop Digital Watermarking (IWWD), pages 30–38, Siena, Italy, Sep 2005.

[14] T. Filler and J. J. Fridrich. Fisher information determines capacity of \( \epsilon \)-secure steganography. In Proc. of 11th International Conference on Information Hiding, pages 31–47, Darmstadt, Germany, 2009.

[15] A. D. Ker. Estimating steganographic Fisher information in real images. Information Hiding, 11th International Conference, 5806:73–88, 2009.

[16] A. D. Ker, P. Bas, R. Bohme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevny. Moving steganography and steganalysis from the laboratory into the real world. In Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security, pages 45–58, New York, NY, USA, 2013. ACM.

[17] S. Craver and J. Yu. Subset selection circumvents the square root law. In Proc. SPIE Media Forensics Security, page 754103, San Jose, CA, 2010.
\[
(\log M_{1n} + \log M_{2n})(1 - \varepsilon_n) - 2 \\
\leq \sum_{i=1}^{n} I(W_1; Y_{i1} | W_2, Y_{i1,i+1}^{n}, S) + I(W_2; Y_2^{i-1}, S) \\
\leq \sum_{i=1}^{n} I(W_1, Y_2^{i-1}; Y_{i1} | W_2, S, Y_{i1,i+1}^{n}) + I(W_2, S, Y_2^{i-1}; Y_{i2}) \\
= \sum_{i=1}^{n} I(W_1, Y_2^{i-1}; Y_{i1} | W_2, S, Y_{i1,i+1}^{n}) + I(W_2, S, Y_2^{i-1}; Y_{i2}) - I(Y_1^{n}_{i+1}; Y_{i2} | W_2, S, Y_2^{i-1}) \\
= \sum_{i=1}^{n} I(W_1; Y_{i1} | W_2, S, Y_{i1,i+1}^{n}, Y_2^{i-1}) + I(Y_2^{i-1}; Y_{i1} | W_2, S, Y_{i1,i+1}^{n}) \\
+ I(W_2, S, Y_2^{i-1}; Y_{i2}) - I(Y_1^{n}_{i+1}; Y_{i2} | W_2, S, Y_2^{i-1}) \\
= \sum_{i=1}^{n} I(W_1; Y_{i1} | U_{2i}) + I(U_2; Y_{i2}) \\
\leq \sum_{i=1}^{n} I(X_i; Y_{i1} | U_{2i}) + I(U_2; Y_{i2}).
\]

[18] M. Tahmasbi and M. R. Bloch. Second-order asymptotics of covert communications over noisy channels. In Proc. IEEE Int. Symp. Inform. Theory, pages 2224–2228, Barcelona, Spain, Jul 2016.

[19] M. Tahmasbi, M. R. Bloch, and V. Y. F. Tan. Error exponent for covert communications over discrete memoryless channels. In Proc. of the IEEE Inform. Th. Workshop, pages 304–308, Kaohsiung, Taiwan, Nov 2017.

[20] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell. Covert communication with noncausal channel-state information at the transmitter. In Proc. IEEE Int. Symp. Inform. Theory, pages 2830–2834, Aachen, Germany, Jun 2017.

[21] L. Wang. Optimal throughput for covert communication over a classical-quantum channel. In Proc. of the IEEE Inform. Th. Workshop, pages 364–368, Cambridge, UK, Sep 2016.

[22] K. S. K. Arumugam and M. R. Bloch. Keyless covert communication over multiple-access channels. In Proc. IEEE Int. Symp. Inform. Theory, pages 2229–2233, Barcelona, Spain, Jul 2016.

[23] K. S. K. Arumugam and M. R. Bloch. Covert communication over broadcast channels. In Proc. of the IEEE Inform. Th. Workshop, pages 299–303, Kaohsiung, Taiwan, Nov 2017.

[24] A. Lapidoth, E. Telatar, and R. Urbanke. On wide-band broadcast channels. IEEE Trans. Inf. Theory, 49(12):3250–3258, Dec 2003.

[25] A. El Gamal. The capacity of a class of broadcast channels. IEEE Trans. Inf. Theory, 25(2):166–169, 1979.

[26] C. Nair. Upper concave envelopes and auxiliary random variables. International Journal of Advances in Engineering Sciences and Applied Mathematics, 5(1):12–20, 2013.

[27] A. Wyner and J. Ziv. A theorem on the entropy of certain binary sequences and applications–I. IEEE Trans. Inf. Theory, 19(6):769–772, 1973.

[28] C. E. Shannon. A mathematical theory of communication. The Bell Systems Technical Journal, 27:379–423, 1948.

[29] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control, 2:101–112, Jun 1959.

[30] T. Cover. Broadcast channels. IEEE Trans. Inf. Theory, 18(1):2–14, 1972.

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

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

[33] M. Tahmasbi and M. R. Bloch. First and second order asymptotics in covert communication with pulse-position modulation. CoRR, abs/1703.01362v2, 2017.

[34] R. G. Gallager. Capacity and coding for degraded broadcast channels. Problems of Information Transmission, 10(3):3–14, 1974.

[35] J. Hou. Coding for relay networks and effective secrecy for wiretap channels. PhD thesis, Technische Universität München, Munich, Germany, 2014.

Vincent Y. F. Tan (S’07-M’11-SM’15) was born in Singapore in 1981. He is currently an Associate Professor in the Department of Electrical and Computer Engineering and the Department of Mathematics at the National University of Singapore (NUS). He received the B.A. and M.Eng. degrees in Electrical and Information Sciences from Cambridge University in 2005 and the Ph.D. degree in Electrical Engineering and Computer Science (EECS) from the Massachusetts Institute of Technology (MIT) in 2011. His research interests include information theory, machine learning, and statistical signal processing.

Dr. Tan received the MIT EECS Jin-Au Kong outstanding doctoral thesis prize in 2011, the NUS Young Investigator Award in 2014, the NUS Engineering Young Researcher Award in 2018, and the Singapore National Research Foundation (NRF) Fellowship (Class of 2018). He is also an IEEE Information Theory Society Distinguished Lecturer for 2018/9. He has authored a research monograph on “Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities” in the Foundations and Trends in Communications and Information Theory Series (Now Publishers). He is currently serving as an Associate Editor of the IEEE Transactions on Signal Processing.

Si-Hyun 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.

Department of Electrical and Computer Engineering