Embedding Covert Information on a Given Broadcast Code

David Kibloff, Samir Perlaza, Ligong Wang

Embedding Covert Information on a Given Broadcast Code

David Kibloff, Samir M. Perlaza, and Ligong Wang

***Abstract*—Given a code used to send a message to two receivers through a degraded discrete memoryless broadcast channel (DM-BC), the sender wishes to alter the codewords to achieve the following goals: (i) the original broadcast communication continues to take place, possibly at the expense of a tolerable increase of the decoding error probability; and (ii) an additional covert message can be transmitted to the stronger receiver such that the weaker receiver cannot detect the existence of this message. The main results are: (a) feasibility of covert communications is proven by using a random coding argument for general DM-BCs; and (b) necessary conditions for establishing covert communications are described and an impossibility (converse) result is presented for a particular class of DM-BCs. Together, these results characterize the asymptotic fundamental limits of covert communications for this particular class of DM-BCs within an arbitrarily small gap.**

***Index Terms*—Covert Communication, Low Probability of Detection, Information-Theoretic Security, Broadcast Channel.**

I. Introduction

Covert communications refer to scenarios in which legitimate parties aim at communicating while keeping an adversary unaware of the existence of the communication. In point-to-point channels, reliable covert communications are subject to a fundamental limit that states that only $O(\sqrt{n})$ bits can be transmitted in n channel uses [1]–[4].

Two different covert communication problems have been studied within the context of broadcast channels [5]–[7]. In [5], the sender tries to send two covert messages to two receivers. In [6] and [7], the sender sends a common non-covert message to both receivers, and tries to simultaneously send a covert message to one of the receivers. That is, the other receiver cannot know whether or not a covert message is being sent.

The current work is related to [6] and [7]. The focus is on the problem of embedding a covert message in a non-covert broadcast code. Some of the main differences between this problem and the one in [6] and [7] are:

- In [6] and [7], the non-covert broadcast code and the covert code are designed together by the transmitter. This potentially allows the transmitter to choose a non-covert code on which it is easy to embed a covert code. Alternatively, the current work assumes that the non-covert code is given and cannot be changed, making the achievability proof more difficult.[1]
- In [6] and [7] there is a separate covertness criterion conditional on every non-covert message. In this work, only one covertness criterion on the overall distribution is adopted. This difference considerably complicates the proof of the converse. In fact, a general proof of the converse using the Kullback-Leibler divergence as the covertness criterion is still an open problem. On the other hand, in this work, the total variation distance is used by adapting techniques developed in [8]. Interestingly, the proof of the converse is shown to be tight for a class of channels satisfying certain symmetry properties.

In a nutshell, it is shown that in the scenario considered in this paper, it is possible to covertly transmit $O(\sqrt{n})$ bits in n channel uses by modifying an existing broadcast code. Moreover, the proposed transmission rate is shown to be asymptotically optimal for a class of discrete memoryless broadcast channels (DM-BCs).

The proofs are omitted in this paper due to a space limitation. Interested readers are referred to [9].

Some notation: The function $Q\colon \mathbb{R} \to [0,1]$ denotes the complementary cumulative distribution function of a standard Gaussian random variable, and $Q^{-1}\colon [0,1] \to \mathbb{R}$ denotes its inverse function. For two probability mass functions P_X and Q_X, the function $\chi_k\left(P_X, Q_X\right)$, with $k \in \mathbb{N}$, is

$$\chi_k\left(P_X, Q_X\right) \triangleq \sum_{x \in \mathcal{X}} \frac{\left(P_X(x) - Q_X(x)\right)^k}{Q_X(x)^{k-1}}. \tag{1}$$

II. System Model

Consider a three-party communication system in which a transmitter simultaneously sends information to two receivers through a noisy communication medium. In this work, the noisy communication medium is described by a product random transformation

$$(\mathcal{X}^n, \mathcal{Y}_1^n \times \mathcal{Y}_2^n, P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}), \tag{2a}$$

where $n \in \mathbb{N}$ is the block-length; the alphabets $\mathcal{X}$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are finite; and $\boldsymbol{Y}_1 = (Y_{1,1}, Y_{1,2}, \ldots, Y_{1,n}) \in \mathcal{Y}_1^n$, $\boldsymbol{Y}_2 = (Y_{2,1}, Y_{2,2}, \ldots, Y_{2,n}) \in \mathcal{Y}_2^n$ and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n) \in \mathcal{X}^n$ are n-dimensional vectors of random variables. In particular, given an input $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the output $(\boldsymbol{y}_1, \boldsymbol{y}_2)$

D. Kibloff and S. M. Perlaza are with the Laboratoire CITI, a joint laboratory between the Institut National de Recherche en Informatique et en Automatique (INRIA), the Université de Lyon and the Institut National de Sciences Apliquées (INSA) de Lyon. 6 Av. des Arts 69621 Villeurbanne, France. (david.kibloff, samir.perlaza@inria.fr)

L. Wang is with the Laboratoire ETIS, a joint laboratory between the Université Paris Seine, Université de Cergy-Pontoise, ENSEA and CNRS. 6, Av. du Ponceau 95014 Cergy-Pontoise, France (e-mail: ligong.wang@ensea.fr).

[1] A technical condition is that the given non-covert code must have a positive error exponent; see (27).

with $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ for all $k \in \{1,2\}$ is observed with probability:

$$P_{\boldsymbol{Y}_1\boldsymbol{Y}_2|\boldsymbol{X}}(\boldsymbol{y}_1,\boldsymbol{y}_2|\boldsymbol{x}) = \prod_{t=1}^{n} P_{Y_1|X}(y_{1,t}|x_t) P_{Y_2|Y_1}(y_{2,t}|y_{1,t}). \quad (2b)$$

That is, the channel is degraded and memoryless.

A. Broadcast Codes

The common message index to be sent from the Transmitter to both receivers is a realization of a random variable W that is uniformly distributed in the set $\mathcal{W} = \{1, 2, \ldots, M\}$, with $M \in \mathbb{N}$. To send a common message index within n channel uses, the Transmitter uses an (n, M, ϵ)-broadcast code.

Definition 1 ((n, M, ϵ)-broadcast code)**.** *Given $M \in \mathbb{N}$, $\epsilon \in [0,1]$ and a block-length $n \in \mathbb{N}$, an (n, M, ϵ)-broadcast code for the random transformation in* (2) *is a system*

$$\Big\{ \Big(\boldsymbol{u}(1), \mathcal{D}_1(1), \mathcal{D}_2(1)\Big), \Big(\boldsymbol{u}(2), \mathcal{D}_1(2), \mathcal{D}_2(2)\Big), \ldots, \Big(\boldsymbol{u}(M), \mathcal{D}_1(M), \mathcal{D}_2(M)\Big) \Big\}, \quad (3)$$

that satisfies for all $(i, j, k) \in \mathcal{W}^2 \times \{1,2\}$, with $i \neq j$:

$$\boldsymbol{u}(i) \triangleq (u_1(i), u_2(i), \ldots, u_n(i)) \in \mathcal{X}^n, \quad (4a)$$

$$\mathcal{D}_k(i) \cap \mathcal{D}_k(j) = \emptyset, \quad (4b)$$

$$\bigcup_{l=1}^{M} \mathcal{D}_k(l) \subseteq \mathcal{Y}_k^n, \quad \textit{and} \quad (4c)$$

$$\frac{1}{M} \sum_{i=1}^{M} \Pr\left[\boldsymbol{Y}_k \in \mathcal{D}_k^{\mathsf{c}}(i) \big| \boldsymbol{X} = \boldsymbol{u}(i)\right] \leqslant \epsilon. \quad (4d)$$

The probability operator in (4d) applies with respect to the marginal $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ of the joint distribution in (2b); and $\mathcal{D}_k^{\mathsf{c}}(i)$ in (4d) represents the complement of $\mathcal{D}_k(i)$ with respect to $\mathcal{Y}_k^n$.

Given a broadcast code represented by the system in (3), the Transmitter uses the codeword $\boldsymbol{u}(i)$ to transmit the message index $i \in \mathcal{W}$. At channel use t, with $t \in \{1, 2, \ldots, n\}$, the Transmitter sends the symbol $u_t(i)$ through the channel. For all $k \in \{1,2\}$, Receiver k observes the output $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$ after n channel uses and determines that the message index i was transmitted if it satisfies the decoding rule:

$$\boldsymbol{y}_k \in \mathcal{D}_k(i). \quad (5)$$

The average decoding error probability associated to the given broadcast code at Receiver k, denoted by λ_k, is given in the left hand-side of (4d).

B. Induced Codes

Let the private message index be represented by a random variable $\hat{W}$, independent of W and uniformly distributed over $\hat{\mathcal{W}} = \{1, 2, \ldots, \hat{M}\}$, with $\hat{M} \in \mathbb{N}$. Assume that a broadcast code denoted by $\mathcal{C}$ is given and is represented by the system in (3). The transmitter uses an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code to transmit both the common and private message indices.

Definition 2 ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code)**.** *Given $\hat{M} \in \mathbb{N}$, $\hat{\epsilon} \in [0,1]$, and an (n, M, ϵ)-broadcast code $\mathcal{C}$ described by* (3)*, an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code is a system*

$$\Big\{ \left(\boldsymbol{v}(1,1), \mathcal{D}_1(1,1), \mathcal{D}_2(1)\right), \left(\boldsymbol{v}(1,2), \mathcal{D}_1(1,2), \mathcal{D}_2(1)\right), \ldots, \Big(\boldsymbol{v}(M,\hat{M}), \mathcal{D}_1(M,\hat{M}), \mathcal{D}_2(M)\Big) \Big\}, \quad (6)$$

that satisfies for all $(i,k,j,l) \in \mathcal{W}^2 \times \hat{\mathcal{W}}^2$, with $(i,j) \neq (k,l)$:

$$\boldsymbol{v}(i,j) \triangleq (v_1(i,j), v_2(i,j), \ldots, v_n(i,j)) \in \mathcal{X}^n, \quad (7a)$$

$$\mathcal{D}_1(i,j) \cap \mathcal{D}_1(k,l) = \emptyset, \quad (7b)$$

$$\bigcup_{p=1}^{M} \bigcup_{q=1}^{\hat{M}} \mathcal{D}_1(p,q) \subseteq \mathcal{Y}_1^n, \quad (7c)$$

$$\frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_1 \in \mathcal{D}_1^{\mathsf{c}}(i,j) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right] \leqslant \hat{\epsilon}, \quad (7d)$$

$$\frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \Pr\left[\boldsymbol{Y}_2 \in \mathcal{D}_2^{\mathsf{c}}(i) | \boldsymbol{X} = \boldsymbol{v}(i,j)\right] \leqslant \hat{\epsilon}. \quad (7e)$$

The probability operators in (7d) and (7e) apply with respect to the conditional marginals $P_{\boldsymbol{Y}_1|\boldsymbol{X}}$ and $P_{\boldsymbol{Y}_2|\boldsymbol{X}}$ of the joint distribution in (2b), respectively. The sets $\mathcal{D}_1^{\mathsf{c}}(i,j)$ and $\mathcal{D}_2^{\mathsf{c}}(i)$ represent the complement of $\mathcal{D}_1(i,j)$ and $\mathcal{D}_2(i)$ with respect to $\mathcal{Y}_1^n$ and $\mathcal{Y}_2^n$, respectively. Given an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code denoted by $\hat{\mathcal{C}}$ and described by (6), the Transmitter uses the codeword $\boldsymbol{v}(i,j)$ to transmit the common message index $i \in \mathcal{W}$ and the private message index $j \in \hat{\mathcal{W}}$. At channel use t, with $t \in \{1, 2, \ldots, n\}$, the Transmitter inputs the symbol $v_t(i,j)$ to the channel. At the end of n channel uses, Receiver k observes the output $\boldsymbol{y}_k = (y_{k,1}, y_{k,2}, \ldots, y_{k,n})$, with $k \in \{1,2\}$. Receiver 1 declares that the pair $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$ was transmitted if (i,j) satisfies the decoding rule:

$$\boldsymbol{y}_1 \in \mathcal{D}_1(i,j). \quad (8)$$

On the other hand, the decoding rule of Receiver 2 remains (5), with $k = 2$; *i.e.*, the same as in the broadcast code $\mathcal{C}$.
The average decoding error probability associated to the induced code $\hat{\mathcal{C}}$ at Receiver k is denoted by $\hat{\lambda}_k$ and given in the left hand-side of (7d) and (7e).

One of the central parameters to characterize the $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code $\hat{\mathcal{C}}$ described by (6) is the number of times a component of a codeword $\boldsymbol{u}(i)$ from $\mathcal{C}$ differs from that of the induced codeword $\boldsymbol{v}(i,j)$ from $\hat{\mathcal{C}}$, with $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$. This quantity, denoted by $\omega(i,j)$, is referred to as the *weight of the codeword* $\boldsymbol{v}(i,j)$ and

$$\omega(i,j) \triangleq \sum_{t=1}^{n} \mathbb{1}_{\{u_t(i) \neq v_t(i,j)\}}. \quad (9)$$

The codes $\mathcal{C}$ and $\hat{\mathcal{C}}$ induce several empirical probability mass functions that are relevant for the analysis of induced codes. These functions are defined hereunder.

Definition 3 (Empirical Probability Distributions)**.** *Given an (n, M, ϵ)-broadcast code $\mathcal{C}$ represented by the system in* (3)*,*

consider an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code $\hat{\mathcal{C}}$ represented by the system in (6)*. For all $(x, \hat{x}) \in \mathcal{X}^2$,*

- *the empirical channel input probability distribution induced by the broadcast code $\mathcal{C}$, denoted by $\bar{P}_X$, is*

$$\bar{P}_X(x) \triangleq \frac{1}{nM} \sum_{i=1}^{M} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}}; \tag{10}$$

- *the empirical joint probability distribution induced by the two codes $\mathcal{C}$ and $\hat{\mathcal{C}}$ on $\mathcal{X}^2$, denoted by $\bar{P}_{X\hat{X}}$, is*

$$\bar{P}_{X\hat{X}}(x, \hat{x}) \triangleq \frac{1}{nM\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{\hat{x}=v_t(i,j)\}};$$

- *the empirical probability with which a symbol x in a codeword from $\mathcal{C}$ is changed into a symbol $\hat{x} \neq x$ in a codeword from $\hat{\mathcal{C}}$, denoted by $\hat{P}_{\hat{X}|X}$, is:*

$$\hat{P}_{\hat{X}|X}(\hat{x}|x) \triangleq \frac{\displaystyle\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{\hat{x}=v_t(i,j)\}} \mathbb{1}_{\{x\neq\hat{x}\}}}{\displaystyle\sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} \sum_{t=1}^{n} \mathbb{1}_{\{x=u_t(i)\}} \mathbb{1}_{\{u_t(i)\neq v_t(i,j)\}}},$$

  and $\operatorname{supp} \hat{P}_{\hat{X}|X=x} = \mathcal{X} \setminus \{x\}$*;*
- *the empirical probability with which a symbol x in a codeword from $\mathcal{C}$ is changed to any other symbol to generate a codeword in $\hat{\mathcal{C}}$, denoted by $\theta(x)$, is*

$$\theta(x) \triangleq 1 - \bar{P}_{\hat{X}|X}(x|x),$$

  where $\bar{P}_{\hat{X}|X}(x|x)$ is such that $\bar{P}_{\hat{X}X}(x, x) = \bar{P}_X(x)\bar{P}_{\hat{X}|X}(x|x)$.

C. Covert Codes

Consider an (n, M, ϵ)-broadcast code described by (3) and denoted by $\mathcal{C}$. Consider also an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code denoted by $\hat{\mathcal{C}}$ and described by (6). For all $k \in \{1, 2\}$, let $Q_{\boldsymbol{Y}_k}$ and $R_{\boldsymbol{Y}_k}$ be the probability mass functions of the channel output vector $\boldsymbol{Y}_k$ when the broadcast code $\mathcal{C}$ is used and when the induced code $\hat{\mathcal{C}}$ is used, respectively.

That is, for all $\boldsymbol{y} \in \mathcal{Y}_k^n$,

$$Q_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M} \sum_{i=1}^{M} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{u}(i)), \text{ and} \tag{11}$$

$$R_{\boldsymbol{Y}_k}(\boldsymbol{y}) \triangleq \frac{1}{M\hat{M}} \sum_{i=1}^{M} \sum_{j=1}^{\hat{M}} P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{v}(i,j)), \tag{12}$$

where $P_{\boldsymbol{Y}_k|\boldsymbol{X}}$ is the marginal obtained from (2b). Using this notation a covert code is defined hereunder.

Definition 4 ($(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code)**.** *Given an (n, M, ϵ)-broadcast code $\mathcal{C}$ described by* (3)*, an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code described by* (6) *is said to be an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code if $\|Q_{\boldsymbol{Y}_2} - R_{\boldsymbol{Y}_2}\|_{\mathrm{TV}} \leqslant \delta$, where $Q_{\boldsymbol{Y}_2}$ and $R_{\boldsymbol{Y}_2}$ are defined in* (11) *and* (12)*, respectively.*

The information that can be covertly transmitted to Receiver 1 using an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code is $\log_2\left(\hat{M}\right)$ bits every n channel uses. Thus, given the broadcast code $\mathcal{C}$, a fundamental limit on the information rate at which information can be covertly transmitted is given by the largest possible $\hat{M}$ for which an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code exists. This notion is formalized by the following definition.

Definition 5 (Largest covert code's size)**.** *Given an (n, M, ϵ)-broadcast code $\mathcal{C}$, the largest covert code's size induced by $\mathcal{C}$, denoted by $\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta)$, is:*

$$\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta) = \max\left\{\hat{M} \in \mathbb{N} : \exists\, (n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)\text{-covert code}\right\}.$$

D. A Class of Discrete Memoryless Broadcast Channels

This section introduces a class of DM-BCs for which the achievability and converse bounds presented later are tight. This class of channels are described by the random transformation in (2) subject to the conditions that for all pairs $(x, x') \in \mathcal{X}^2$ with $x \neq x'$,

$$\chi_2\left(P_{Y_2|X=x}, P_{Y_2|X=x'}\right) = d, \tag{13}$$

$$D\left(P_{Y_1|X=x} || P_{Y_1|X=x'}\right) = \ell, \tag{14}$$

where $(d, \ell) \in \mathbb{R}_+^2$. A simple example of a channel satisfying (13) and (14) is the binary symmetric channel. We present another example below.

Example 1. *Consider the random transformation in* (2) *such that $\mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, 2\}$, and such that for all $(x, x') \in \mathcal{X}^2$ with $x \neq x'$, the conditional probability distributions $P_{Y_1|X}$ and $P_{Y_2|Y_1}$ respectively satisfy:*

$$P_{Y_1|X}(x|x) = 1 - 2P_{Y_1|X}(x'|x) = 1 - 2p_1, \text{ and} \tag{15}$$

$$P_{Y_2|Y_1}(x|x) = 1 - 2P_{Y_2|Y_1}(x'|x) = 1 - 2p_2, \tag{16}$$

with $(p_1, p_2) \in]0, \frac{1}{3}[^2$.

III. Achievability Results

This section presents conditions under which an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code exists, where $\mathcal{C}$ is a given (n, M, ϵ)-broadcast code. This result is obtained using a random coding argument and is described in Section III-A. Building upon the existence of an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code in the finite block-length regime, a lower bound on $\hat{M}^*(n, \mathcal{C}, \hat{\epsilon}, \delta)$ is then established in the asymptotic regime where n grows large. This is presented in Section III-B.

A. Finite Block-Length Analysis

Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$ for the random transformation in (2) described by the system in (3). Consider also the parameters $\hat{M} \in \mathbb{N}$; $K \in [0, \sqrt{n}]$ and a conditional probability distribution $\tilde{P}_{\hat{X}|X}$ such that, for all $x \in \mathcal{X}$,

$$\operatorname{supp} \tilde{P}_{\hat{X}|X=x} \subseteq \mathcal{X} \setminus \{x\}. \tag{17}$$

Using the parameters K and $\tilde{P}_{\hat{X}|X}$, let $P_{\hat{X}|X}$ be a conditional probability distribution such that for all $(x, \hat{x}) \in \mathcal{X}^2$,

$$P_{\hat{X}|X}(\hat{x}|x) \triangleq (1 - \theta)\, \mathbb{1}_{\{x=\hat{x}\}} + \theta \tilde{P}_{\hat{X}|X}(\hat{x}|x), \tag{18}$$

with

$$\theta \triangleq \frac{K}{\sqrt{n}}. \tag{19}$$

For all $i \in \{1, 2, \ldots, M\}$, generate $\hat{M}$ codewords

$$\boldsymbol{v}(i,1), \boldsymbol{v}(i,2), \ldots, \boldsymbol{v}(i,\hat{M}) \tag{20}$$

to form the codebook of an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon})$-induced code. For all $j \in \{1, 2, \ldots, \hat{M}\}$, the codeword $\boldsymbol{v}(i,j)$ is the realization of a random variable following the probability distribution $P_{\hat{\boldsymbol{X}}|\boldsymbol{X}=\boldsymbol{u}(i)}$ such that for all $\hat{\boldsymbol{x}} \in \mathcal{X}^n$,

$$P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}\left(\hat{\boldsymbol{x}}|\boldsymbol{u}(i)\right) \triangleq \prod_{t=1}^{n} P_{\hat{X}|X}\left(\hat{x}_t|u_t(i)\right), \tag{21}$$

where $\boldsymbol{u}(1)$, $\boldsymbol{u}(2)$, …, $\boldsymbol{u}(M)$ are the codewords of the given broadcast code $\mathcal{C}$.

For all $(\boldsymbol{x}, \hat{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{X}^{2n} \times \mathcal{Y}_k^n$ and for all $k \in \{1, 2\}$, let $\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x})$ be defined by

$$\imath_k(\hat{\boldsymbol{x}}; \boldsymbol{y}|\boldsymbol{x}) \triangleq \log_2\left(\frac{P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})}{\sum\limits_{\boldsymbol{x}' \in \mathcal{X}^n} P_{\hat{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{x}'|\boldsymbol{x}) P_{\boldsymbol{Y}_k|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}')}\right).$$

Receiver 1 uses the decoding sets

$$\mathcal{D}_1(i,j) = \Big\{\boldsymbol{y} \in \mathcal{D}_1(i) : \imath_1\Big(\boldsymbol{v}(i,j), \boldsymbol{y}|\boldsymbol{u}(i)\Big) \geqslant n\eta\Big\} \setminus \bigcup_{k<j} \mathcal{D}_1(i,k), \tag{22}$$

with $(i,j) \in \mathcal{W} \times \hat{\mathcal{W}}$, where $\eta \in \mathbb{R}$ is a design parameter. Recall that Receiver 2 uses the decoding sets corresponding to the given broadcast code $\mathcal{C}$, which are given by (5) with $k = 2$. Note that the codewords in (20), the decoding sets in (22), and the decoding sets in (5) for $k = 2$ form an induced code.

Define the following notation:

$$\bar{D}(\tilde{P}_{\hat{X}|X}) \triangleq \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \bar{P}_X(x) \tilde{P}_{\hat{X}|X}(\hat{x}|x) D\big(P_{Y_1|X=\hat{x}} || P_{Y_1|X=x}\big), \tag{23}$$

and for all $k \in \{1, 2\}$ and for all pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}_k$, let $\tilde{R}_{Y_k|X}(y|x)$ be the distribution

$$\tilde{R}_{Y_k|X}(y|x) \triangleq \sum_{\hat{x} \in \mathcal{X}} \tilde{P}_{\hat{X}|X}(\hat{x}|x) P_{Y_k|X}(y|\hat{x}), \tag{24}$$

and

$$\bar{\chi}_2(\tilde{R}_{Y_k|X}, P_{Y_k|X}) \triangleq \sum_{x \in \mathcal{X}} \bar{P}_X(x) \sum_{y \in \mathcal{Y}_k} \frac{\big(\tilde{R}_{Y_k|X}(y|x) - P_{Y_k|X}(y|x)\big)^2}{P_{Y_k|X}(y|x)}. \tag{25}$$

Using the above random construction, the existence of a covert code can be proved, as given by the following proposition.

Proposition 1. *Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$ for the random transformation in* (2)*. Then, there always exists an $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code that satisfies*

$$\frac{\log_2\left(\hat{M}\right)}{\sqrt{n}} \geqslant \max_{\tilde{P}_{\hat{X}|X}} \frac{2(1-\xi)\bar{D}(\tilde{P}_{\hat{X}|X}) Q^{-1}\Big(\frac{1-\delta-\epsilon-\hat{\epsilon}+\sqrt{c_n}-\frac{c}{\sqrt{n}}}{2}\Big)}{\sqrt{\bar{\chi}_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})}}, \tag{26}$$

where c is a positive constant, c_n is such that $\lim_{n\to\infty} c_n = 0$, and $\xi \in]0, 1[$ can be chosen arbitrarily small.

B. Asymptotic Analysis

In the regime in which the block-length n grows large, Proposition 1 leads to the following result.

Theorem 1. *Consider a sequence $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, …, of (n, M_n, ϵ_n)-broadcast codes for the random transformation in* (2)*, with $n \in \{1, 2, \ldots\}$ and*

$$\epsilon_n \leqslant \exp(-\zeta n), \tag{27}$$

for some fixed positive real ζ. Then, there always exists a sequence of $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert codes with $\lim_{n\to\infty} \hat{\epsilon}_n = 0$, such that

$$\liminf_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} \geqslant \max_{\tilde{P}_{\hat{X}|X}} \frac{2(1-\xi)\bar{D}(\tilde{P}_{\hat{X}|X})}{\sqrt{\bar{\chi}_2(\tilde{R}_{Y_2|X=x}, P_{Y_2|X=x})}} Q^{-1}\Big(\frac{1-\delta}{2}\Big), \tag{28}$$

with $\xi \in]0, 1[$ arbitrarily small.

Theorem 1 implies that it is possible to covertly transmit $O(\sqrt{n})$ bits in n channel uses by modifying an existing broadcast code. In the next section, the bound (28) is shown to be tight for the class of DM-BCs described in Section II-D.

IV. Converse Results

This section introduces some necessary conditions on the parameters of any $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code in the finite block-length regime. Then an asymptotic upper bound is presented for the case in which the random transformation in (2) satisfies (13) and (14).

A. Finite Block-Length Analysis

Using Fano's inequality [10], the following proposition presents for every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code $\hat{\mathcal{C}}$ an upper bound on $\log_2\left(\hat{M}\right)$ in terms of the empirical probability mass functions induced by both the original code $\mathcal{C}$ and the covert code $\hat{\mathcal{C}}$ (Definition 3).

Proposition 2. *Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$, described by the system in* (3)*, for the random transformation in* (2)*. Then, every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code satisfies*

$$\log_2\left(\hat{M}\right) \leqslant \frac{1}{1-\hat{\epsilon}}\Bigg(1 + n \sum_{x \in \mathcal{X}} \sum_{\hat{x} \in \mathcal{X}} \bar{P}_X(x)\theta(x)\hat{P}_{\hat{X}|X}(\hat{x}|x) \cdot D\big(P_{Y_1|X=\hat{x}} || P_{Y_1|X=x}\big) + \bar{P}_X(x)\frac{\theta(x)^3}{6}\chi_3(\hat{R}_{Y_1|X=x}, P_{Y_1|X=x})\Bigg). \tag{29}$$

Given a covert code, a covert sub-code can be obtained by choosing the codewords whose weight is bounded. More importantly, the cardinality of the set of upper-bounded-weight codewords can be lower-bounded. The next proposition is inspired by [8].

Proposition 3. *Let $\eta > 0$ be arbitrarily small. Consider an (n, M, ϵ)-broadcast code $\mathcal{C}$, described by the system in* (3)*, for the random transformation in* (2)*. Assume that the random transformation in* (2) *satisfies* (13) *and* (14)*. Then, every $(n, \mathcal{C}, \hat{M}, \hat{\epsilon}, \delta)$-covert code described by the system in* (6) *can be formed by two sub-codes: one sub-code whose codewords are in the set*

$$\tilde{\mathcal{W}} = \Bigg\{ \boldsymbol{v}(i,j) : \omega(i,j) < 2\sqrt{\frac{n}{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), 1 \leqslant i \leqslant M, \text{ and } 1 \leqslant j \leqslant \hat{M} \Bigg\}; \tag{30}$$

and another sub-code whose codewords are in the set

$$\tilde{\mathcal{W}}^{\mathrm{c}} = \Bigg\{ \boldsymbol{v}(i,j) : \omega(i,j) \geqslant 2\sqrt{\frac{n}{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), 1 \leqslant i \leqslant M, \text{ and } 1 \leqslant j \leqslant \hat{M} \Bigg\}. \tag{31}$$

Moreover, $\left|\tilde{\mathcal{W}}\right| > M\hat{M}\left(\frac{\eta}{2} - \frac{c}{\sqrt{n}} - \epsilon - \hat{\epsilon}\right)$, where $c > 0$ is a constant.

B. Asymptotic Analysis

The following theorem introduces the main result of this section.

Theorem 2. *Consider a sequence $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \ldots$, of (n, M_n, ϵ_n)-broadcast codes for the random transformation in* (2)*, with $\lim_{n\to\infty} \epsilon_n = 0$. Assume that the random transformation in* (2) *satisfies* (13) *and* (14)*. Then, for any sequence $\hat{\mathcal{C}}_1$, $\hat{\mathcal{C}}_2$, $\hat{\mathcal{C}}_3$, $\ldots$ of $(n, \mathcal{C}_n, \hat{M}_n, \hat{\epsilon}_n, \delta)$-covert codes with $\lim_{n\to\infty} \hat{\epsilon}_n = 0$, it holds that*

$$\limsup_{n\to\infty} \frac{\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)}{\sqrt{n}} < \frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta-\eta}{2}\right), \tag{32}$$

where d and ℓ are given in (13) *and* (14)*, respectively, and $\eta > 0$ can be arbitrarily small.*

Note that, for channels belonging to the class in Section II-D, the right-hand side of (28) simplifies to

$$(1-\xi)\frac{2\ell}{\sqrt{d}} Q^{-1}\left(\frac{1-\delta}{2}\right), \tag{33}$$

with $\xi \in]0, 1[$ arbitrarily small. Recalling that η in (32) can be chosen to be arbitrarily close to zero, it follows that, for such channels, the asymptotic bounds in Theorems 1 and 2 are tight; i.e., (33) gives the optimal scaling constant for $\log_2\left(\hat{M}_n^*(n, \mathcal{C}_n, \hat{\epsilon}_n, \delta)\right)$ with respect to $\sqrt{n}$.

V. Discussion

Theorem 1 introduces a general achievability result that reveals that it is possible to covertly transmit $O(\sqrt{n})$ bits in n channel uses. Theorem 2 introduces an impossibility (converse) result that is particular to the class of DM-BCs described in Section II-D. Together, they provide a characterization of the asymptotic growth rate for the number of bits that can be covertly transmitted in DM-BCs of this class.

Note that the scheme proposed here might not be optimal for asymmetric channels. Indeed, if the channel is asymmetric, it might be easier to replace a particular symbol $x \in \mathcal{X}$ in the original codeword with another one in the covert codeword. Therefore, in general one should allow the parameter θ to depend on the symbol x that is replaced.

So far, a tight converse for general DM-BCs, i.e., those that do not necessarily satisfy the conditions described in Section II-D, is still an open problem. An interesting question is whether the total variation distance used in the current work can be replaced by the Kullback-Leibler divergence.

Finally, it is interesting to highlight that the problem introduced in this paper is an instance of a more general problem. In multi-user channels, broadcast codes can be altered to perform other functionalities, e.g., simultaneous energy and information transmission to an energy harvester, physical-layer secrecy, *etc*.

References

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal on Selected Areas in Communications*, 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. of IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2945–2949.

[3] M. Bloch, "Covert communications over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

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

[5] V. Y. F. Tan and S. Lee, "Time-division transmission is optimal for covert communication over broadcast channels," 2017. [Online]. Available: http://arxiv.org/abs/1710.09754

[6] K. S. K. Arumugam and M. R. Bloch, "Covert communication over broadcast channels," in *Proc. of IEEE Information Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 299–303.

[7] K. S. K. Arumugam and M. R. Bloch, "Embedding covert information in broadcast communications," *IEEE Transactions on Information Forensics and Security*, 2019.

[8] M. Tahmasbi and M. R. Bloch, "Second-order asymptotics in covert communication," 2017. [Online]. Available: http://arxiv.org/abs/1703.01362

[9] D. Kibloff, S. M. Perlaza, and L. Wang, "Broadcast codes can be enhanced to perform covert communications," INRIA Grenoble - Rhône-Alpes, Tech. Rep. 9249, Jan. 2019.

[10] R. Fano, *Transmission of Information: A Statistical Theory of Communication*, 1st ed. MIT Press, 1961.