Covert Communication over Noisy Channels: A Resolvability Perspective

Matthieu R. Bloch

Abstract

We consider the situation in which a transmitter attempts to communicate reliably over a noisy memoryless channel, while simultaneously ensuring covertness (low probability of detection) with respect to a warden, who observes the signals through another noisy memoryless channel. We develop three coding schemes inspired by principles from source and channel resolvability, which generalize and extend prior work in two directions. First, we show that, irrespective of the quality of the channels, it is possible to communicate \(O(\sqrt{n})\) reliable and covert bits over \(n\) channel uses if the transmitter and the receiver share a key of size \(O(\sqrt{n})\); this improves upon earlier results requiring a key of size \(O(\sqrt{n \log n})\) bits. Second, we show that, if the receiver’s channel is “better” than the warden’s channel in a sense that we precise, it is possible to communicate \(O(\sqrt{n})\) reliable and covert bits over \(n\) channel uses without secret key; this generalizes earlier results established for binary symmetric channels. The main technical problem that we address is how to develop concentration inequalities for “low-weight” sequences; the crux of our approach is to define suitably modified typical sets that are amenable to concentration inequalities.

I. INTRODUCTION

The benefits offered by ubiquitous communication networks are now mitigated by the relative ease with which malicious users can interfere or tamper with sensitive data. The past decade has thus witnessed a growing concern for the issues of privacy, confidentiality, and integrity of communications. In many instances, users in a communication network find themselves in a position in which they wish to communicate without being detected by others. Such situations include fairly innocuous scenarios of dynamic spectrum access in wireless channels, in which secondary users attempt to communicate without being detected by primary users. A perhaps more adversarial example is a situation in which a user wishes to convey information covertly, either to maintain his privacy, avoid attacks, or escape the attention of regulatory entities monitoring the network.

Motivated by these challenges, [1], [2] have established the first characterization of the maximum throughput at which two users communicate reliably over a noisy channel while guaranteeing a low probability of interception from a warden, who observes the transmitted signal though another noisy channel. In particular, it has been shown that arbitrarily low probability of detection over pure loss quantum channels, thermal noise quantum channels, and classical Gaussian channels, is possible as long as one communicates at most \(O(\sqrt{n})\) bits over \(n\) uses of the channel; one notable characteristic of the proposed communication schemes is to require a secret key between the legitimate users of size \(O(\sqrt{n \log n})\). Such results may be viewed as the counterparts of the “square root law” of steganography [3] when the message is embedded in a cover with zero mean. The results of [1], [2] have been extended in several directions, in particular by showing that arbitrarily probability detection is possible without secret-key when all users are connected by Binary Symmetric Channels (BSCs) and provided the warden’s BSC noise is larger than legitimate users’ BSC noise. Others extensions have attempted to identify scenarios in which the “square root law” may be beaten, which includes situations in which the channel statistics are imperfectly known [4], [5], [6], or when the warden has uncertainty about the time of communication [7]. The ideas underlying the key-less coding scheme are also connected to those developed for “stealth” and channel resolvability in the context of wiretap channels [8], [9]. Tutorial presentations and discussions of these results may be found in [10], [11].

In the remaining of the paper, we use the terminology “covert communication” as a synonym for low-probability of detection [12], [2], [1], deniability [13], [4], and undetectable communication [5], [6]. The main conceptual contribution of this technical report is to revisit the problem of covert communication from the perspective of resolvability [14], [15]. This conceptual connection allows us to treat continuous and discrete channels alike, and to establish the following three technical results that improve on and extend earlier work.

- We develop an alternative coding scheme to [1] that shows that \(O(\sqrt{n})\) reliable and covert bits may be communicated over \(n\) channel uses with \(O(\sqrt{n \log_2 n})\) bits of secret key in a universal manner; this is essentially a variation [1] with a technical refinement (Theorem [1]).
If the warden’s channel statistics are known, we show that \(O(\sqrt{n})\) reliable covert bits may be communicated over \(n\) channel uses with only \(O(\sqrt{n})\) bits of secret key (Theorem 2).

If the legitimate user’s channel is “better” than the warden’s channel, in a sense that is made precise in Section IV, \(O(\sqrt{n})\) reliable covert bits may be communicated over \(n\) channel uses without secret key: in particular, this generalizes [13] to all Discrete Memoryless Channels (DMCs) and Gaussian channels (Theorem 3).

The underlying technical problem that we solve is how to develop random coding arguments for “low weight” channel are denoted \(W\) through another memoryless channel (over a memoryless channel (including Gaussian channels).

For covert communication without secret key, Section V presents several applications and extension of the results, establishing the possibility of covert communication assisted by secret key, while Section IV establishes conditions for covert communication without secret key. Section V presents several applications and extension of the results, including Gaussian channels.

II. COVERT COMMUNICATION OVER NOISY CHANNELS

We consider the situation illustrated in Fig. 1 in which two legitimate users, Alice and Bob, attempt to communicate over a memoryless channel \((\mathcal{X}, W_{Y|X}, \mathcal{Z})\) without being detected by a warden, Willie, who observes the signals through another memoryless channel \((\mathcal{X}, W_{Z|X}, \mathcal{Z})\). The transition probabilities corresponding to \(n\) uses of the channel are denoted \(W_{Y|X}^n \triangleq \prod_{i=1}^n W_{Y|X}\) and \(W_{Z|X}^n \triangleq \prod_{i=1}^n W_{Z|X}\). We also make the following assumptions.

- There exists an innocent symbol \(x_0 \in \mathcal{X}\) that corresponds to the input to the channel when no communication takes place. In such a case, the distributions induced by \(x_0\) at the output of the two memoryless channels are
  \[
  P_0 \triangleq W_{Y|X=0} \quad \text{and} \quad Q_0 \triangleq W_{Z|X=0}.
  \]

- There exists another symbol \(x_1 \in \mathcal{X}\) with \(x_1 \neq x_0\), and we define the distributions induced by \(x_1\) at the output of the memoryless channels
  \[
  P_1 \triangleq W_{Y|X=x_1} \quad \text{and} \quad Q_1 \triangleq W_{Z|X=x_1}.
  \]

- \(P_1\) is absolutely continuous with respect to \(P_0\), denoted \(P_1 \ll P_0\), and \(Q_1 \ll Q_0\).

Although these assumptions restrict the class of channels considered, they are nevertheless satisfied for large classes of channels; for instance, for Additive White Gaussian Noise (AWGN) channels, \(x_0 = 0\) is the natural choice of the innocent symbols, and the absolute continuity requirements are satisfied.

![Fig. 1. Covert communication over noisy channels.](image)

Formally, the objective of Alice is to transmit a message \(W\) uniformly distributed in \([1, M]\) by encoding them into codewords \(X = (X_1, \ldots, X_n)\) of \(n\) symbols with the help of a secret key \(S\) uniformly distributed in \([1, K]\). Upon observing a noisy version \(Y = (Y_1, \ldots, Y_n)\) of \(X\) and knowing \(S\), Bob’s objective is to form a reliable estimate \(\hat{W}\) of \(W\), with reliability measured by the average probability of error \(P(W \neq \hat{W})\). In contrast, Willie’s goal is to perform a statistical test on his observation \(Z = (Z_1, \ldots, Z_n)\) to decide whether the legitimate users communicate (hypothesis \(H_0\)) or not (hypothesis \(H_1\)). The probability of type I error (rejecting \(H_0\) when true)
is denoted \( \alpha \), while the probability of type II error (accepting \( H_0 \) when wrong) is denoted \( \beta \). It is possible for the warden to design blind tests that ignore his channel observations and that achieve any pair \((\alpha, \beta)\) such that \( \alpha + \beta = 1 \). Therefore, the objective of covert communication is to guarantee that the warden’s best statistical test yields a trade-off between \( \alpha \) and \( \beta \) that is not much better than that of a blind test. Specifically, let \( Q_0^n \triangleq \prod_{i=1}^n Q_0 \) be the product distribution that is expected by the warden when no communication happens, and let \( \hat{Q}^n \) be the distribution expected when communication takes place. It can be shown [16] that Willie’s optimal hypothesis test yields the tradeoff \( \alpha + \beta = 1 - \mathbb{V}(Q_0^n, \hat{Q}^n) \), where \( \mathbb{V}(. , .) \) is the variational distance between distributions. Therefore, achieving covert communication amounts to ensuring that \( \mathbb{V}(Q_0^n, \hat{Q}^n) \) is negligible.

Consequently, we aim to establish scalings of \( \log_2 M \) and \( \log_2 K \) with \( n \) for which there exists covert communication schemes with

\[
\lim_{n \to \infty} \mathbb{P}(W \neq \hat{W}) = 0 \quad \text{and} \quad \lim_{n \to \infty} \mathbb{V}(Q_0^n, \hat{Q}^n) = 0.
\]

**Remark.** For simplicity, we assume that communication always takes place, and that the goal is to prevent Willie from proving it with a statistical test. The case in which communication might not take place can be modeled as in [13], and merely requires a minor modification of Bob’s decoder.

A. Distances between distributions

We start by recalling definitions and a few known results about distances between distributions. Let, \( P_1 \) and \( P_2 \) denote two measures with respective densities \( f_1 \) and \( f_2 \) with respect to (w.r.t.) to some common measure \( \mu \). We consider the following distances.

\[
\mathbb{V}(P_1 - P_2) \triangleq \frac{1}{2} \| P_1 - P_2 \|_1 = \frac{1}{2} \int |f_1 - f_2| \, d\mu \quad \text{(variational distance)}
\]

\[
H(P_1, P_2) \triangleq \left( \int \left( \sqrt{f_1} - \sqrt{f_2} \right)^2 \, d\mu \right)^{\frac{1}{2}} \quad \text{(Hellinger distance)},
\]

\[
\mathbb{D}(P_2 \| P_1) \triangleq \int \frac{f_2}{f_1} \, dP_1 \quad \text{(Kullback-Leibler divergence)}
\]

\[
\chi(P_2 \| P_1) \triangleq \left( \int \frac{f_2}{f_1} - 1 \right)^2 \, dP_1 \quad \text{(\( \chi \) distance)}
\]

The distances \( \mathbb{D}(P_2 \| P_1) \) and \( \chi(P_2 \| P_1) \) remain finite if \( P_2 \ll P_1 \). In the remainder of the paper, we use the standard notation if discrete random variables and probability mass functions for simplicity, although all results hold for continuous random variables with probability densities. The following relations between the distances are known.

**Lemma 1** (Adapted from [17]).

- \( \| P_1 - P_2 \| \leq H(P_1, P_2) \).
- If \( P_2 \ll P_1 \), then \( H(P_1, P_2) \leq \chi(P_2 \| P_1) \) and \( H^2(P_1, P_2) \leq \mathbb{D}(P_2 \| P_1) \).
- If \( \{P_i\}_{i \in [1,n]} \) and \( \{P'_i\}_{i \in [1,n]} \) are probability measures defined on the same subspace, then

\[
\left\| \prod_{i=1}^n P_i - \prod_{i=1}^n P'_i \right\| \leq \sum_{i=1}^n \| P_i - P'_i \|,
\]

\[
H \left( \prod_{i=1}^n P_i, \prod_{i=1}^n P'_i \right) \leq \left( \sum_{i=1}^n H(P_i, P'_i)^2 \right)^{\frac{1}{2}}.
\]

Note that (9) is a tighter upper bound than (8) thanks to the presence of a square root; unfortunately, (9) is rarely amenable to further analysis because of the intricate expression of the Hellinger distance. We circumvent this issue by upper bounding the Hellinger distance by the \( \chi \) distance. Although the bounds are looser, they end up having much more expressive power.
B. Covert processes

We now go back to the setting of Fig. 1. For \( n \in \mathbb{N}^* \), let \( \alpha_n \in (0, 1) \) be of the form \( \alpha_n = \frac{\omega_n}{\sqrt{n}} \) with
\[
\lim_{n \to \infty} \omega_n = 0 \quad \text{and} \quad \lim_{n \to \infty} \omega_n \sqrt{n} = \infty.
\tag{10}
\]
Define the input distribution \( \Pi_{\alpha_n} \) on \( \{x_0, x_1\} \) such that \( \Pi_{\alpha_n}(x_1) = 1 - \Pi_{\alpha_n}(x_0) = \alpha_n \), as well as the corresponding output distributions
\[
Q_{\alpha_n}(z) = \sum_x W_{Z|X}(z|x) \Pi_{\alpha_n}(x) = W_{Z|X}(z|x_1)\alpha_n + W_{Z|X}(z|x_0)(1 - \alpha_n),
\tag{11}
\]
\[
P_{\alpha_n}(y) = \sum_x W_{Y|X}(y|x) \Pi_{\alpha_n}(x) = W_{Y|X}(y|x_1)\alpha_n + W_{Y|X}(y|x_0)(1 - \alpha_n).
\tag{12}
\]
Also define the product distributions
\[
\Pi_{\alpha_n}^n = \prod_{i=1}^{n} \Pi_{\alpha_n}, \quad Q_{\alpha_n}^n = \prod_{i=1}^{n} Q_{\alpha_n}, \quad \text{and} \quad P_{\alpha_n}^n = \prod_{i=1}^{n} P_{\alpha_n}.
\tag{13}
\]
Note that \( Q_1 \ll Q_0 \) implies \( Q_{\alpha_n} \ll Q_0 \) and that \( P_1 \ll P_0 \) implies \( P_{\alpha_n} \ll P_0 \). We then have the following result, which proof may be found in Appendix A.

**Lemma 2.** Let \( Q_{\alpha_n} \) and \( Q_0 \) be defined as per (I) and (II). Then,
\[
\mathcal{V}(Q_{\alpha_n}^n, Q_0^n) \leq \sqrt{n} \alpha_n \lambda(Q_1||Q_0).
\]
Consider the joint random variables \( (X, Z) \in \{x_0, x_1\} \times Z \) with distribution \( W_{Z|X}(z|x) \Pi_{\alpha_n}(x) \). Then,
\begin{itemize}
  \item \( \mathbb{I}(X; Z) \geq \alpha_n \mathbb{D}(Q_1||Q_0) - \alpha^2_n \lambda(Q_1||Q_0)^2 \).
  \item \( \mathbb{I}(X; Z) \leq \alpha_n \mathbb{D}(Q_1||Q_0) - \alpha^2_n \mathcal{V}(Q_1, Q_0)^2 \).
\end{itemize}

**Remark.** The bounds on \( \mathbb{I}(X; Z) \) given by Lemma 2 are not tight since \( \mathcal{V}(Q_1, Q_0) \leq H(Q_0, Q_1) \leq \lambda(Q_1||Q_0) \) by Lemma 7 and one easily exhibits distributions for which the inequality is strict. Nevertheless, the first order in \( \alpha_n \) is tight, which is all we use in the remainder of the paper. In addition, the result allows us to circumvent rather painful Taylor series of \( \mathbb{I}(X; Z) \) in \( \alpha_n \).

For the specific choice of \( \alpha_n \) in (10), notice that \( \lim_{n \to \infty} \mathcal{V}(Q_{\alpha_n}^n, Q_0^n) = 0 \), so that \( Q_{\alpha_n}^n \) becomes indistinguishable from \( Q_0^n \); therefore, we call the process \( Q_{\alpha_n}^n \) a “covert stochastic process.” In addition, the realizations of the input process \( \Pi_{\alpha_n} \) contain an average of \( \omega_n \sqrt{n} \) \( x_1 \) symbols, which grows to infinity with \( n \); this opens the possibility of embedding information symbols in the channel input that would remain covert. Essentially, the result of Lemma 2 formalizes the intuition that as long as the number of \( x_1 \) symbols transmitted in a sequence of \( n \) symbols does not exceed \( O(\sqrt{n}) \), then the change in the distribution perceived the warden is indistinguishable from the statistical noise. The fact that a stochastic process with a non-trivial number of \( x_1 \) symbols may induce an undetectable covert stochastic process at the output of a noisy channel, suggests a generic principle for the design of stealth communication schemes, which we formulate as follows.

**Covert communication schemes should attempt to simulate a covert stochastic process** \( Q_{\alpha_n}^n \).

The covert communication schemes developed in Section III and Section IV correspond to different applications of this principle.

C. Technical digression: concentration inequalities with low-weight sequences

One of the technical challenges faced when trying to deal with stochastic processes such as \( \Pi_{\alpha_n} \) in (13), is that the concentration inequalities traditionally used to develop information-theoretic results do not readily apply here. To be more concrete, consider the joint random variables \( (X, Z) \in \mathcal{X}^n \times \mathcal{Z}^n \) with the product distribution \( \prod_{i=1}^{n} W_{Z|X}(z_i|x_i) \Pi_{\alpha_n}(x_i) \); define the mutual information random variable \( (15) \)
\[
\log \frac{W_{Z|X}(Z)}{Q_{\alpha_n}^n(Z)} = \log \frac{W_{Z|X}(Z)}{Q_{\alpha_n}^n(Z)}, \tag{14}
\]
whose average is the average mutual information $\mathbb{I}(X; Z) = n \mathbb{I}(X; Z)$. Assuming for simplicity that the range of
\[
\log \frac{W_{Z|X}(Z|X)}{Q_{\alpha_n}(Z)}
\] is a finite interval of length $\eta > 0$\(^1\). Hoeffding’s inequality states that for any $\mu > 0$
\[
\mathbb{P}\left( \left| \log \frac{W_{Z|X}(Z|X)}{Q_{\alpha_n}(Z)} - \mathbb{I}(X; Z) \right| \geq \mu \mathbb{I}(X; Z) \right) \leq 2 \exp \left( \frac{-2n\mu^2 \mathbb{I}(X; Z)^2}{\eta^2} \right). \tag{15}
\]

Unfortunately, this upper does not vanish because of the specific scaling of $\mathbb{I}(X; Z)$ with $n$ given in Lemma 2. The problem finds its roots in the fact that the sequences $X$ have “low weight,” i.e., the number of $x_1$ symbols is on average on the order of $\omega_n \sqrt{n}$, which is sub-linear in $n$.

There are, however, some concentration inequalities that are still useful and that will be exploited in virtually all subsequent proofs. For instance, consider a binary random sequence $S \in \{0, 1\}^n$ with a product distribution $\prod_{i=1}^n P_S$ such that $P_S(1) = 1 - P_S(0) = \omega_n \sqrt{n}$. The sequence $S$ is of low average weight $\omega_n \sqrt{n}$, but the application of a Chernoff bound [18] Exercise 2.10 yields for any $\mu \in (0, 1)$
\[
\mathbb{P}\left( \left| \sum_{i=1}^n S_i - \omega_n \sqrt{n} \right| > \mu \omega_n \sqrt{n} \right) \leq 2 \exp \left( -\frac{-\mu^2 \omega_n \sqrt{n}}{3} \right), \tag{16}
\]
which vanishes with our choice of $\omega_n$. The difference between (16) and (15) may be intuitively understood as follows. The number of terms contributing to $\sum_{i=1}^n S_i$ in (16) is on average $\omega_n \sqrt{n}$ because most terms are zero. In contrast, all the terms in $\sum_{i=1}^n \log \frac{W_{Z|X}(Z|X)}{Q_{\alpha_n}(Z)}$ are potential contributors to the sum in (15); the concentration inequality (15) fails because the individual contributions of the terms in the sum are too small.

III. KEY-ASSISTED COVERT COMMUNICATION SCHEMES

In this section, we develop two “key-assisted” covert communication schemes that operate with a secret key $S$ and allow the transmission of $O(\sqrt{n})$ bits over $n$ channel uses. The general architecture of the schemes is similar to [11], but the exact operation differs and the required key size required varies depending on the assumptions regarding the knowledge of the warden’s channel statistics.

![Key-assisted covert communication scheme](image)

As illustrated in Fig. 2, the schemes operate according to the following general principle.
1) Alice and Bob split the secret key $S$ into two keys $\tilde{S} \in [1, \tilde{K}]$ and $\hat{S} \in [1, \hat{K}]$ such that $\tilde{K} \hat{K} = K$.
2) Alice and Bob spread the secret key $\tilde{S}$ into a length $n$ sequence $\tilde{X} \in \{x_0, x_1\}^n$.
3) Alice encodes the message $W$ into a length $n'$ binary codeword $B \in \{0, 1\}^{n'}$, where $n'$ will be specified later.
4) Alice transmits information by modulating the symbols of $\tilde{X}$ in the position $i$ for which $\tilde{X}_i = x_1$, resulting in a transmitted sequence $\tilde{X}$. Formally, consider realizations $\tilde{x}$, $b$, and $\tilde{s}$, define
\[
\text{supp}(\tilde{x}) \triangleq \{ i \in [1, n'] : \tilde{x}_i \neq x_0 \}, \tag{17}
\]
\(^1\)This holds if the channel $(X, W_{Z|X}, Z)$ is a fully connected DMC, such as a BSC.
and let \( \{i_j\} \) with \( j \in [1, \text{supp}(\tilde{x})] \) be the positions for which \( \tilde{x}_{i_j} \neq x_0 \). The symbols of the modulated sequence \( \tilde{x} \) are defined as

\[
\tilde{x}_i = \begin{cases} 
    x_{b_j \oplus \delta_j}, & \text{if } \exists j \in \left[1, \min(\text{supp}(\tilde{x}), n')\right] \text{ such that } i = i_j \\
    x_{\delta_j}, & \text{if } \exists j \in \left[\min(\text{supp}(\tilde{x}), n'), \text{supp}(\tilde{x})\right] \text{ such that } i = i_j \\
    x_0 & \text{otherwise.}
\end{cases}
\tag{18}
\]

Effectively, the modulated sequence \( \tilde{X} \) is obtained by transmitting the sequence \( \tilde{X} \) through a memoryless \( Z \)-channel, in which the \( x_0 \) symbol is unaffected and the \( x_1 \) symbol is flipped to the \( x_0 \) symbol with probability \( \frac{1}{2} \). We denote the transition probability of this \( Z \) channel by \( W_{\tilde{X}|\tilde{X}} \), and we let

\[
W_{Z|X}(z|x) \triangleq \sum_{\hat{x}} W_{Z|X}(z|\hat{x}) W_{\tilde{X}|\tilde{X}}(\hat{x}|x).
\tag{19}
\]

5) Upon observing the channel output \( Y \), Bob uses his knowledge of \( \tilde{X} \) to create a sequence \( \hat{Y} = (Y_1, \ldots, Y_{\text{supp}(\tilde{x})}) \).

If \( \text{supp}(\tilde{X}) < n' \), Bob declares an error; otherwise, it attempts to decode \( \hat{Y} \) with \( \hat{S} \) to form an estimate \( \hat{W} \) of \( W \).

A. Universal covert communication with \( O(\sqrt{n} \log n) \) secret key bits

Following the principle outlined in Section II-B, we first attempt to simulate the process \( Q_{\alpha_n} \) by simulating the process \( \Pi_{\alpha_n} \) defined as per [13] at the input of the channel. Specifically, the secret key \( S \) is encoded into a sequence \( \tilde{X} \in \{x_0, x_1\}^n \) such that the distribution \( P_{\tilde{X}} \) of \( \tilde{X} \) is close in variational distance to \( \Pi_{\alpha_n} \). The following theorem characterizes the performance of this covert communication scheme.

**Theorem 1.** Let \( C \) be the capacity of the main channel with inputs restricted to \( \{x_0, x_1\} \). For any \( \alpha > 0 \), there exist \( \alpha_1, \alpha_2 > 0 \) depending on \( W_{Y|X} \) but not on \( W_{Z|X} \), and a covert communication scheme as in Fig. 2 such that, for \( n \) large enough:

- \( \log_2 M = (1 - \alpha)C\omega_n \sqrt{n} \);
- \( \log_2 K = (1 + \alpha)\omega_n \sqrt{n} \log_2 n \);
- \( P(W \neq \hat{W}) \leq e^{-\alpha_1 \omega_n \sqrt{n}} \);
- \( \nu(Q^n, Q_0^n) \leq e^{-\alpha_2 \omega_n \sqrt{n}} + \omega_n (Q_1||Q_0) \).

This scheme is universal w.r.t. the warden’s channel, in the sense that \( \nu(Q^n, Q_0^n) \) is bounded for \( n \) large enough as soon as \( \chi(Q_1||Q_0) \) is bounded, irrespective of the exact statistics \( W_{Z|X} \).

**Remark.** With some extra work, one may prove that the key \( \tilde{S} \) is not necessary. Specifically, one can develop a random coding argument that includes the random generation of the code used by the legitimate users, and establish similar results without relying on a key \( \tilde{S} \). We omit the proof, which is slightly more involved but does not affect the scaling in \( n \).

**Proof:** The proof of Theorem 1 consists in showing the existence of a deterministic encoder to generate \( \tilde{X} \) from the key \( \tilde{S} \), and the existence of a codebook with length approximately \( \omega_n \sqrt{n} \) that modulates \( \tilde{X} \) into \( \tilde{X} \).

a) **Existence of spreading code:** Let \( \tilde{K} \in \mathbb{N}^s \), \( \epsilon \in (0, 1) \), and define the set

\[
\mathcal{T}_\epsilon \triangleq \{ x \in \mathcal{X}^n : (1 - \epsilon)\omega_n \sqrt{n} \leq \text{supp}(x) \leq (1 + \epsilon)\omega_n \sqrt{n} \}.
\tag{20}
\]

Generate \( \tilde{K} \) codewords \( \tilde{x}_i \in \{x_0, x_1\}^n \) of length \( n \) independently according to the distribution

\[
\Pi_{\lambda_n}^n(x) \triangleq \frac{1}{\lambda_n} \Pi_{\alpha_n}^n(x) \quad \text{with } \lambda_n \triangleq \mathbb{P}_{\Pi_{\alpha_n}^n}(X \in \mathcal{T}_\epsilon^n).
\tag{21}
\]

Using a Chernoff bound, we have

\[
1 - \lambda_n = \mathbb{P}_{\Pi_{\alpha_n}^n}(X \notin \mathcal{T}_\epsilon^n) \leq 2 \exp \left( - \frac{\epsilon^2}{3} \omega_n \sqrt{n} \right),
\tag{22}
\]
and one may check that
\[
\mathbb{V} (\Pi_{\alpha_n}^n, \Pi_{\alpha_n, \epsilon}^n) \leq 1 - \lambda_n + \left( \frac{1}{\lambda_n} - 1 \right) \leq 6 \exp \left( -\frac{\epsilon^2}{3} \omega_n \sqrt{n} \right) \text{ for } n \text{ large enough.} \tag{23}
\]

Finally, define the output distribution corresponding to \( \Pi_{\alpha_n, \epsilon}^n \) as
\[
Q_{\alpha_n, \epsilon}^n \triangleq \sum_x W_{Z|X}(z|x) \Pi_{\alpha_n, \epsilon}^n (x). \tag{24}
\]

The encoder spreads a secret key \( \tilde{s} \in [1, \tilde{K}] \) into a sequence \( \tilde{x} \in \{ x_0, x_1 \}^n \) according to the map \([1, \tilde{K}] \to \{ x_0, x_1 \}^n : \tilde{s} \mapsto \tilde{x}_s \). The resulting spread sequence distribution is then
\[
P_{\tilde{X}} (x) = \frac{1}{\tilde{K}} 1 \{ x = \tilde{x}_s \}. \tag{25}
\]

Our objective is to show that for suitably large \( \tilde{K} \), the spread sequence distribution \( P_{\tilde{X}} \) is close in variational distance to the product distribution \( \Pi_{\alpha_n}^n \). This is actually a variation of source resolvability [14], which we detail to carefully handle the dependence of \( \Pi_{\alpha_n}^n \) on \( \alpha_n \). As shown in Appendix B, one may establish the following result.

Lemma 3. For any \( \gamma > 0 \), the average of \( \mathbb{V} (P_{\tilde{X}}, \Pi_{\alpha_n, \epsilon}^n) \) over the random code generation satisfies
\[
\mathbb{E} \left( \mathbb{V} (P_{\tilde{X}}, \Pi_{\alpha_n, \epsilon}^n) \right) \leq \frac{1}{\lambda_n} \mathbb{P}_{\Pi_{\alpha_n}^n} \left( \text{supp}(X) \geq \gamma + n \frac{\ln(1 - \alpha_n)}{\ln \frac{1 - \alpha_n}{\alpha_n}} \right) + \frac{1}{2} \sqrt{\frac{\epsilon \gamma}{\lambda_n \tilde{K}}}. \tag{26}
\]

For any \( \mu > 0 \), by choosing
\[
\gamma = (1 + \mu) \omega_n \sqrt{n} \ln \left( \frac{1}{\alpha_n} - 1 \right) - n \ln(1 - \alpha_n) \tag{27}
\]
and noticing that \( \text{supp}(X) = \sum_{i=1}^n 1 \{ X_i = 1 \} \) with \( \mathbb{E}_{\Pi_{\alpha_n}^n} (\text{supp}(X)) = \omega_n \sqrt{n} \), we obtain with a Chernoff bound
\[
\mathbb{P}_{\Pi_{\alpha_n}^n} \left( \text{supp}(X) \geq \gamma + \frac{n \ln(1 - \alpha_n)}{\ln \frac{1 - \alpha_n}{\alpha_n}} \right) = \mathbb{P}_{\Pi_{\alpha_n}^n} \left( \text{supp}(X) > (1 + \mu) \omega_n \sqrt{n} \right) \leq \exp \left( -\frac{\mu^2}{3} \omega_n \sqrt{n} \right). \tag{28}
\]

With \( \alpha_n = \frac{\omega}{\sqrt{n}} \), notice that
\[
\gamma = (1 + \mu) \omega_n \sqrt{n} \ln \left( \frac{1}{\alpha_n} - 1 \right) - n \ln(1 - \alpha_n) \leq (1 + \mu) \omega_n \sqrt{n} \left( \ln \sqrt{n} - \ln \omega_n \right) + \frac{n \omega_n}{\sqrt{n} - \omega_n} \tag{29}
\]
where we have used the inequality \( \ln(1 + x) \geq \frac{x}{1+x} \) for \( x \in ]-1; +\infty[ \). For \( n \) large enough, we also have \( \ln \sqrt{n} - \ln \omega_n < \ln n \) by (10) and \( \sqrt{n} - \omega_n \geq \frac{\sqrt{n}}{\mu \ln n} \), so that \( \gamma \leq (1 + 2\mu) \omega_n \sqrt{n} \ln n \). Hence, choosing
\[
\ln \tilde{K} = (1 + 2\delta)(1 + 2\mu) \omega_n \sqrt{n} \ln n \quad \text{with any} \quad \delta > 0, \tag{30}
\]
we obtain for \( n \) large enough
\[
\mathbb{E} \left( \mathbb{V} (P_{\tilde{X}}, \Pi_{\alpha_n, \epsilon}^n) \right) \leq \frac{1}{\lambda_n} \exp \left( -\frac{\mu^2}{3} \omega_n \sqrt{n} \right) + \frac{1}{2 \sqrt{\lambda_n}} \exp \left( -\delta(1 + 2\mu) \omega_n \sqrt{n} \ln n \right)
\]
\[
\leq 2 \exp \left( -\rho \omega_n \sqrt{n} \right) \quad \text{with some appropriately defined} \quad \rho > 0. \tag{31}
\]

In particular, there exists a specific code for which \( \mathbb{V} (P_{\tilde{X}}, \Pi_{\alpha_n, \epsilon}^n) < 4 \exp \left( -\rho \omega_n \sqrt{n} \right) \).
b) Effect of modulation: Irrespective of the error-control code used to encode $W$, modulation requires the use of at most
\[
\ln \hat{K} = (1 + \epsilon) \omega_n \sqrt{n} \text{ bits}
\]
by the constraint imposed in (20). When presenting the distribution $\Pi_{\alpha_n}$ at the input of the $Z$-channel $W_{X|\hat{X}}$ induced by the modulation, one may check that the corresponding distribution at the output of the $Z$-channel is $\Pi_{\beta_n}$ with $\beta_n \triangleq \alpha_n/2$. Consequently, the distribution $\hat{Q}^n$ induced by the coding scheme satisfies,
\[
\forall(\hat{Q}^n, Q_0^n) \leq \forall(\hat{Q}^n, Q_{\beta_n}^n) + \forall(Q_{\beta_n}^n, Q_0^n) \\
\leq \forall(P_{\hat{X}}, \Pi_{\alpha_n}^n) + \forall(Q_{\beta_n}^n, Q_0^n) \\
\leq \forall(P_{\hat{X}}, \Pi_{\alpha_n}^n) + \forall(\Pi_{\alpha_n}^n, \Pi_{\alpha_n}^n) + \forall(Q_{\beta_n}^n, Q_0^n) \\
\leq 4 \exp\left(-\rho \omega_n \sqrt{n}\right) + 6 \exp\left(-\frac{\epsilon^2}{3} \omega_n \sqrt{n}\right) + \omega_n \epsilon (Q_1 \| Q_0),
\]
where we have used the data processing inequality of variational distance in (34) to show $\forall(\hat{Q}^n, Q_{\beta_n}^n) \leq \forall(P_{\hat{X}}, \Pi_{\alpha_n}^n)$.  

c) Reliability: We conclude the proof by showing how one may encode the messages $W$ into codewords $B$. Assume that the main channel has capacity $C$ nats when inputs are restricted to the set $\{x_0, x_1\}$. Standard arguments [19] show that, for any $\delta > 0$, there exists a binary code of length $(1 - \delta) \omega_n \sqrt{n}$, such that one may transmit
\[
\ln M = (1 - \delta)(1 - \epsilon)C \omega_n \sqrt{n} \text{ nats}
\]
with probability of error
\[
P_{\text{err}} \leq \exp\left(-\delta' (1 - \epsilon) \omega_n \sqrt{n}\right),
\]
where $\delta' > 0$ depend on $\delta$ and $W_{Y|X}$.  
Combining the choice of $\ln \hat{K}$, $\ln \hat{K}$, and $\ln M$, in (30), (32), (37), with the bounds obtained in (36) and (38), one may then find appropriate constants as promised in the statement of the theorem.  

**Corollary 1.** Let $C$ be as in Theorem 1 and let $\epsilon > 0$. There exists universal covert communication schemes with length $n$ such that
\[
\begin{align*}
\log_2 M &= O(\sqrt{n}), \\
\log_2 \hat{K} &= O(\sqrt{n} \log_2 n), \\
\lim_{n \to \infty} \forall(\hat{Q}^n, Q_0^n) &= \epsilon \text{ and } \lim_{n \to \infty} \mathbb{P}(W \neq \hat{W}) = 0.
\end{align*}
\]

**Proof:** Apply Theorem 1 replacing $\omega_n$ by $\epsilon' \chi(Q_1 \| Q_0)$ with $0 < \epsilon' < \epsilon$.  

Corollary 1 is essentially a refinement of [11, Theorem 1.2]. It differs from [11] by ensuring a bound on the maximum key size instead on the average key size, although it is still $O(\sqrt{n} \log_2 n)$. Despite the similarity of the results, Theorem 1 relies on more sophisticated resolvability techniques [15], whose usefulness will become apparent in Section III-B and Section IV. The underlying covert communication schemes are also somewhat different, as the key $K$ may be viewed in our scheme as the seed to generate a “spreading sequence.”

**B. Covert communication with $O(\sqrt{n})$ secret key bits**

A weakness of the covert communication scheme analyzed in Theorem 1 is that it requires a key size on the order of $\omega_n \sqrt{n} \log_2 n$ bits to transmit on the order of $\omega_n \sqrt{n}$ bits. In a practical implementation of the proposed coding scheme, the key is likely to stem from a pseudo-random number generator, which opens the proposed scheme to attacks that get particularly detrimental as the required key is long. Fortunately, we show next how the scheme may be suitably modified to use a key size on the order of $\omega_n \sqrt{n}$ bits. The idea behind the improvement is to use the key $\hat{S}$ to directly simulate the covert process $Q_{\alpha_n}^n$ defined as per (13), without simulating the process $\Pi_{\alpha_n}^n$. The architecture remains identical to that in Fig. 2 and only the encoding of $\hat{S}$ is modified. Conceptually, the idea is to
rely on channel resolvability in place of source resolvability, but the precise analysis require some care because of the “low weight” nature of the process \( \Pi_{\alpha_n} \).

**Theorem 2.** Let \( C \) be the capacity of the main channel with inputs restricted to \( \{x_0, x_1\} \) and assume that the random variable \( \ln \frac{Q_1(Z)}{Q_0(Z)} \) for \( Z \sim Q_1 \) is sub-Gaussian. For any \( \alpha > 0 \), there exist \( \alpha_1, \alpha_2 > 0 \) depending on \( \alpha \), \( W_{Y|X}, W_{Z|X} \), and a covert communication scheme as in Fig. 2 such that, for \( n \) large enough:

- \( \log_2 M = (1 - \alpha)C\omega_n\sqrt{n} \)
- \( \log_2 K = (1 + \alpha)\left(1 + \frac{\mathbb{E}(Q_1||Q_0)}{2}\right)\omega_n\sqrt{n} \)
- \( \mathbb{P}(W \neq \hat{W}) \leq e^{-\alpha_1\omega_n\sqrt{n}} \)
- \( \mathbb{V}(\hat{Q}^n, Q^n_0) \leq e^{-\alpha_2\omega_n\sqrt{n}} + \omega_n\epsilon (Q_1||Q_0) \)

**Remark.** A closer inspection of the proof of Theorem 2 shows that the sub-Gaussian condition may be replaced by any other condition that guarantees a concentration result for the sum of \( n \) independent and identically distributed (i.i.d.) copies of \( \ln \frac{Q_1(Z)}{Q_0(Z)} \). In particular, the concentration follows directly from Hoeffding’s inequality in the case of DMCs.

**Corollary 2.** Let \( C \) be as in Theorem 2 and let \( \epsilon > 0 \). If the random variable \( \ln \frac{Q_1(Z)}{Q_0(Z)} \) for \( Z \sim Q_1 \) is sub-Gaussian, there exists covert communication schemes with length \( n \) such that

- \( \log_2 M = O(\sqrt{n}) \)
- \( \log_2 K = O(\sqrt{n}) \)
- \( \lim_{n \to \infty} \mathbb{V}(\hat{Q}^n, Q^n_0) \leq \epsilon \) and \( \lim_{n \to \infty} \mathbb{P}(W \neq \hat{W}) = 0. \)

**Proof:** We will show that there exists a deterministic encoder to generate \( \hat{X} \) from the key \( \hat{S} \), such that the distribution \( \hat{Q}^n \) induced by \( P_{\hat{X}} \) and modulation is close to the covert process \( Q^n_{\beta_n} \) with \( \beta_n \triangleq \frac{\alpha}{2} \) in variational distance. By not asking \( P_{\hat{X}} \) to be close to \( \Pi^n_{\alpha_n} \) as in the proof of Theorem 1 the required key size is reduced. The modulation of \( \hat{X} \) into \( \hat{X} \) remains identical to what was done earlier.

**d) Existence of spreading code:** Let \( \hat{K} \in \mathbb{N}^* \). Define \( \hat{T}^n, \hat{\Pi}_{\alpha_n, \epsilon}, \lambda_n, \) and \( \hat{Q}^n_{\alpha_n, \epsilon} \), as per (20), (21), and (24), respectively. Generate \( \hat{K} \) codewords \( \hat{x}_i \in \{x_0, x_1\}^n \) independently according to the distribution \( \hat{\Pi}^n_{\alpha_n, \epsilon} \). The encoder generates a sequence \( \hat{x} \) from a secret key \( s \in [1, \hat{K}] \) according to the map \( [1, \hat{K}] \to \{x_0, x_1\}^n : \hat{s} \mapsto \hat{x}_s \). The distribution induced by the encoder at the output of the warden’s channel is

\[
\hat{Q}^n(z) \triangleq \frac{1}{\hat{K}} \sum_{i=1}^{\hat{K}} W^n_{Z|\hat{X}}(z|\hat{x}_i) \quad \text{with} \quad W^n_{Z|\hat{X}}(z|\hat{x}) \triangleq \sum_{\tilde{x}} W^n_{Z|X}(z|\tilde{x})W^n_{X|\hat{X}}(\tilde{x}|\hat{x}).
\]

Define

\[
\hat{Q}_0(z) \triangleq W^n_{Z|\hat{X}}(z|x_0) = Q_0(z) \quad \text{and} \quad \hat{Q}_1(z) \triangleq W^n_{Z|\hat{X}}(z|x_1) = \frac{1}{2}Q_0(z) + \frac{1}{2}Q_1(z)
\]

and note that

\[
\mathbb{D}(\hat{Q}_1||\hat{Q}_0) \leq \frac{1}{2}\mathbb{D}(Q_1||Q_0) \quad \text{and} \quad \chi(\hat{Q}_1||\hat{Q}_0) = \frac{1}{2}\chi(Q_1||Q_0).
\]

As shown in Appendix C one may establish the following.

**Lemma 4.** Define \( \hat{Q}^n_{\alpha_n, \epsilon}(z) \triangleq \sum_{x} W^n_{Z|\hat{X}}(z|x)\hat{\Pi}^n_{\alpha_n, \epsilon}(x) \). For any \( \gamma > 0 \), the average of \( \mathbb{V}(\hat{Q}^n, \hat{Q}^n_{\alpha_n, \epsilon}) \) over the random code generation satisfies

\[
\mathbb{E}\left(\mathbb{V}(\hat{Q}^n, \hat{Q}^n_{\alpha_n, \epsilon})\right) \leq \frac{1}{\lambda_n}\mathbb{E}_{W^n_{Z|\hat{X}}}\left[\sum_{i=1}^{n} \ln \frac{W^n_{Z|\hat{X}}(Z_i|X_i)}{\hat{Q}_0(Z_i)} > \gamma\right] + \frac{1}{2}\sqrt{\frac{2\gamma}{\lambda_n\hat{K}}}.
\]

**Remark.** Although Lemma 4 may be viewed as an instance of channel resolvability [14], additional work is required to handle the dependence of \( \hat{\Pi}^n_{\alpha_n} \) on \( \alpha_n \), which decays faster than \( 1/\sqrt{n} \). In particular, the typical set used in standard proofs (Chernoff bounds, etc.) does not seem to be amenable to standard concentration inequalities because the number of bits communicated is on the order of \( \omega_n\sqrt{n} \) over \( n \) channel uses. Our approach to circumvent the
problem was inspired by an astute technique proposed in [13] to “concentrate” the sum of \( n \) i.i.d. random variables over a sum of size \( \sqrt{n} \) terms.

If \( X_i = x_0, Z_i \) is distributed according to \( \widetilde{Q}_0 \) and \( \ln \frac{W_{z_i|x}(Z_i|0)}{Q_0(Z_i)} = 0 \); in contrast, if \( X_i = x_1, Z_i \) is distributed according to \( \widetilde{Q}_1 \) and \( \ln \frac{W_{z_i|x}(Z_i|1)}{Q_0(Z_i)} = \ln \frac{\tilde{Q}_1(Z_i)}{Q_0(Z_i)} \). Consequently, although the first term on the right-hand side of (42) contains a sum of \( n \) terms, only those for which \( X_i = x_1 \) contribute to it. Therefore, we introduce the random variable \( L \triangleq \supp(\mathbf{X}) = \sum_{i=1}^n 1 \{ \tilde{X}_i \neq x_0 \} \), and for \( \mu, \nu > 0 \) we define

\[
\gamma \triangleq (1 + \mu)(1 + \nu) \mathbb{D}(\widetilde{Q}_1 \| \widetilde{Q}_0) \omega_n \sqrt{n} \quad \text{and} \quad \mathcal{D}_\mu^n \triangleq \{ \ell \in \mathbb{N}^* : |\ell - \omega_n \sqrt{n}| < \mu \omega_n \sqrt{n} \}. \tag{43}
\]

Then,

\[
P_{W_{z|x} \Pi_n^n} \left( \sum_{i=1}^n \ln \frac{W_{z|x}(Z_i|X_i)}{Q_0(Z_i)} > \gamma \right) \leq \sum_{\ell \in \mathcal{D}_\mu^n} P(L = \ell) P_{\widetilde{Q}_1} \left( \sum_{i=1}^\ell \ln \frac{\tilde{Q}_1(Z_i)}{Q_0(Z_i)} > \gamma \right) + P(L \notin \mathcal{D}_\mu^n), \tag{44}
\]

with

\[
P(L \notin \mathcal{D}_\mu^n) = P(|L - \omega_n \sqrt{n}| \leq \mu \omega_n \sqrt{n}) \leq 2 \exp \left( -\frac{\mu^2 \omega_n \sqrt{n}}{3} \right). \tag{45}
\]

For \( \ell \in \mathcal{D}_\mu^n \), note that

\[
(1 + \mu)(1 + \nu) \omega_n \sqrt{n} - \ell > (1 + \nu) \ell - \ell = \nu \ell,
\]

so that

\[
P_{\widetilde{Q}_1} \left( \sum_{i=1}^\ell \ln \frac{\tilde{Q}_1(Z_i)}{Q_0(Z_i)} > \gamma \right) = P_{\widetilde{Q}_1} \left( \sum_{i=1}^\ell \ln \frac{\tilde{Q}_1(Z_i)}{Q_0(Z_i)} - \ell \mathbb{D}(\tilde{Q}_1 \| \tilde{Q}_0) \right) \geq \mathbb{D}(\tilde{Q}_1 \| \tilde{Q}_0) \ell \tag{47}
\]

\[
\leq A \exp(-a \ell) \quad \text{for some constants} \ A, a > 0 \tag{48}
\]

\[
\leq A \exp(-a(1 - \mu) \omega_n \sqrt{n}), \tag{50}
\]

where we have use the assumption that \( \ln \frac{\tilde{Q}_1(Z_i)}{Q_0(Z_i)} \) is sub-Gaussian in the last inequality to invoke a Chernoff bound. Combining, the inequalities above and choosing

\[
\ln \bar{K} = (1 + 2\delta)(1 + \mu)(1 + \nu) \frac{\mathbb{D}(\tilde{Q}_1 \| \tilde{Q}_0)}{2} \omega_n \sqrt{n} \geq (1 + 2\delta)(1 + \mu)(1 + \nu) \mathbb{D}(\tilde{Q}_1 \| \tilde{Q}_0) \omega_n \sqrt{n}, \tag{51}
\]

we obtain

\[
\mathbb{E} \left( \mathbb{V}(\tilde{Q}_n^n, \tilde{Q}_{\beta_n, \epsilon}^n) \right) \leq \frac{1}{\lambda_n} \left( A \exp(-a(1 - \mu) \omega_n \sqrt{n}) + 2 \exp \left( -\frac{\mu^2 \omega_n \sqrt{n}}{3} \right) \right) + \frac{1}{2} \exp \left( -\delta(1 + \mu)(1 + \nu) \mathbb{D}(\tilde{Q}_1 \| \tilde{Q}_0) \omega_n \sqrt{n} \right) \leq \frac{1}{\lambda_n} \exp(-\rho_2 \omega_n \sqrt{n}) \quad \text{for some appropriate choice of} \ \rho_2 > 0. \tag{52}
\]

In particular, for \( n \) large enough, there exists a specific code for which \( \mathbb{V}(\tilde{Q}_n^n, \tilde{Q}_{\beta_n, \epsilon}^n) \leq 2 \exp(-\rho_2 \omega_n \sqrt{n}) \).
\textbf{e) Modulation and reliability:} The modulation and reliability analysis follow exactly the steps in the proof of Theorem 1. In particular, modulation requires at most
\begin{equation}
\ln K = (1 + \epsilon)\omega_n\sqrt{n} \text{ nats} \tag{54}
\end{equation}
and for any \(\delta > 0\), one may find a code to reliability transmit
\begin{equation}
\ln M = (1 - \delta)(1 - \epsilon)C\omega_n\sqrt{n} \text{ nats}. \tag{55}
\end{equation}
Finally, note that for \(n\) large enough
\begin{align}
\mathbb{V}(\hat{Q}_n^\top, Q_0^n) &\leq \mathbb{V}(\hat{Q}_n^\top, Q_{\beta_n}^\top) + \mathbb{V}(Q_0^n, Q_{\beta_n}^n) \\
&\leq \mathbb{V}(\hat{Q}_n^\top, Q_{\beta_n,\epsilon}^\top) + \mathbb{V}(Q_{\beta_n}^n, Q_0^n) \\
&\leq \mathbb{V}(\hat{Q}_n^\top, Q_{\beta_n,\epsilon}^\top) + \mathbb{V}(\Pi_{\alpha_n,\epsilon}^\top, \Pi_{\alpha_n}^\top, Q_{\beta_n}^n, Q_0^n) \\
&\leq 2\exp\left(-\rho_2\omega_n\sqrt{n}\right) + 6\exp\left(-\frac{\epsilon^2}{3}\omega_n\sqrt{n}\right) + \frac{\omega_n}{2}\chi(Q_1\|Q_0). \tag{59}
\end{align}
Combing the choice of \(\ln K, \ln K\), and \(\ln M\), in (54), (59), (55), with the bounds obtained in (59) and (38), one may then find appropriate constants as promised in the statement of the theorem.

\section{IV. Key-less covert communications}

We now go back to the architecture illustrated in Fig. 1 in which we consider the possibility of achieving covert \textit{without} any shared secret key between the legitimate users. The objective is to design

\textbf{Theorem 3.} Assume that \(\mathbb{D}(P_1 || P_0) > \mathbb{D}(Q_1 || Q_0)\), the random variables \(\ln \frac{Q_1(Z)}{Q_0(Z)}\) with \(Z \sim Q_1\) and \(\ln \frac{P_1(Y)}{P_0(Y)}\) with \(Y \sim P_1\) are sub-Gaussian, and \(\sum_y \frac{P_1(y)^2}{P_0(y)} < \infty\). Then, for any \(\alpha > 0\), there exists \(\alpha_1, \alpha_2 > 0\) depending on \(\alpha\), \(W_{Y|X}\), and \(W_{Z|X}\), and a key-less covert communication scheme, such that

\begin{itemize}
\item \(\log_2 M = (1 + \alpha)\mathbb{D}(P_1 || P_0)\omega_n\sqrt{n}\) bits,
\item \(\mathbb{P}\left(W \neq \hat{W}\right) \leq e^{-\alpha_1\omega_n\sqrt{n}}\),
\item \(\mathbb{V}(\hat{Q}_n^\top, Q_0^n) \leq e^{-\alpha_2\omega_n\sqrt{n}} + \omega_n\chi(Q_1\|Q_0)\).
\end{itemize}

\textbf{Corollary 3.} Let \(\epsilon > 0\). Under the same conditions as in Theorem 3 there exist covert communication schemes with length \(n\) such that

\begin{itemize}
\item \(\log_2 M = O(\sqrt{n})\),
\item \(\lim_{n \to \infty} \mathbb{V}(\hat{Q}_n^\top, Q_0^n) \leq \epsilon\) and \(\lim_{n \to \infty} \mathbb{P}\left(W \neq \hat{W}\right) = 0\).
\end{itemize}

\textit{Proof:} The proof of Theorem 3 is essentially a random coding argument for channel reliability \(\mathbb{D}(P_1 || P_0)\) and channel resolvability \(\mathbb{D}(Q_1 || Q_0)\); however, because the number of bits communicated is on the order of \(\omega_n\sqrt{n}\) over \(n\) channel uses, traditional concentration inequalities (Chernoff bounds, etc.) do not seem to directly apply. As for the proof of Theorem 2, the crucial idea to circumvent this technical issue is to use suitably modified typical sets.

\textbf{f) Random codebook generation:} Let \(M \in \mathbb{N}^*\). Generate \(M\) codewords \(x_i\) independently according to the product distribution \(\Pi_{\alpha_n}^\top\). Define the set
\begin{equation}
\mathcal{A}_n^\gamma \triangleq \left\{ (x, y) \in \mathcal{X}^n \times \mathcal{Y}^n : \ln \frac{W_{Y|X}^n(y|x)}{P_0^n(y)} \geq \gamma \right\} \tag{60}
\end{equation}
where \(\gamma > 0\) will be determined later. The encoder simply consists in mapping \(i \mapsto x_i\). The decoder consists in mapping \(y\) to \(i \in [1, M]\), if \(x_i\) is the unique codeword such that \((x_i, y) \in \mathcal{A}_n^\gamma\); otherwise, an error is declared.

\textsuperscript{2}The traditional typical set for decoding is similar to \(\mathcal{A}_n^\gamma\) but with \(P_{\alpha_n}^n(y)\) in place of \(P_0^n(y)\) \(\mathbb{D}(P_1 || P_0)\).
g) Channel reliability analysis: According to the definition of the encoder/decoder above, the probability of error averaged of the random codebook is

\[ \mathbb{E}(P_e) = \mathbb{E} \left( \sum_{y} \sum_{i=1}^{M} \frac{1}{M} W_{n}^{n}(y|X_{i})1 \left\{ (X_{1}, y) \notin A_{n} \right\} \right) \]

\[ = \mathbb{E} \left( \sum_{y} W_{n}^{n}(y|X_{1})1 \left\{ (X_{1}, y) \notin A_{n} \right\} \right) + \sum_{j \neq i} \mathbb{E} \left( \sum_{y} W_{n}^{n}(y|X_{1})1 \left\{ (X_{j}, y) \notin A_{n} \right\} \right) \]

\[ \leq \mathbb{E} \left( \sum_{y} W_{n}^{n}(y|X_{1})1 \left\{ (X_{1}, y) \notin A_{n} \right\} \right) + \sum_{j \neq i} \mathbb{E} \left( \sum_{y} W_{n}^{n}(y|X_{1})1 \left\{ (X_{j}, y) \notin A_{n} \right\} \right) \] (61)

We start by analyzing the last term on the right-hand side of (61). For any \( j \neq 1 \),

\[ \mathbb{E} \left( \sum_{y} W_{n}^{n}(y|X_{1})1 \left\{ (X_{j}, y) \in A_{n} \right\} \right) = \sum_{x} \sum_{y} P_{n}(y) \Pi_{n}(x) \left\{ (x, y) \in A_{n} \right\} \]

\[ = \sum_{x} \sum_{y} P_{n}(y) \Pi_{n}(x) \frac{P_{0}(y)}{P_{0}(y)} \left\{ (x, y) \in A_{n} \right\} \]

\[ \leq \sum_{x} \sum_{y} W_{n}^{n}(y|X|x) e^{-\gamma} \Pi_{n}(x) \frac{P_{n}(y)}{P_{0}(y)} \left\{ (x, y) \in A_{n} \right\} \]

\[ \leq e^{-\gamma} \mathbb{E}_{P_{0}}^{n} \left( \frac{P_{n}(Y)}{P_{0}(Y)} \right) \] (65)

Since \( P_{n}^{n} \) and \( Q_{0}^{n} \) are product distributions, we have

\[ \mathbb{E}_{P_{0}}^{n} \left( \frac{P_{n}(Y)}{P_{0}(Y)} \right) = \left( \mathbb{E}_{P_{0}} \left( \frac{P_{n}(Y)}{P_{0}(Y)} \right) \right)^{n} \] (66)

Next, note that

\[ \mathbb{E}_{P_{0}} \left( \frac{P_{n}(Y)}{P_{0}(Y)} \right) = \mathbb{E}_{P_{0}} \left( 1 - \alpha_{n} + \alpha_{n} \frac{P_{1}(Y)}{P_{0}(Y)} \right) \]

\[ = 1 - \alpha_{n} + \alpha_{n} \left( \sum_{y} \left( (1 - \alpha_{n}) \frac{P_{0}(y)}{P_{0}(y)} + \alpha_{n} \frac{P_{1}(y)}{P_{0}(y)} \right) \right) \]

\[ = 1 - \alpha_{n} + \alpha_{n} \left( 1 - \alpha_{n} + \alpha_{n} \sum_{y} \frac{P_{1}(y)}{P_{0}(y)} \right) \]

\[ = 1 + \alpha^{2}_{n}(\xi - 1), \text{ with } \xi \triangleq \sum_{y} \frac{P_{1}(y)}{P_{0}(y)}. \] (70)

Consequently,

\[ \mathbb{E}_{P_{0}}^{n} \left( \frac{P_{n}(Y)}{P_{0}(Y)} \right) = (1 + \alpha^{2}_{n}(\xi - 1))^{n} = \exp \left( n \ln (1 + \alpha^{2}_{n}(\xi - 1)) \right) \leq \exp (n\alpha^{2}_{n}(\xi - 1)) = \exp \left( \omega^{2}_{n}(\xi - 1) \right). \] (71)

Hence, we obtain

\[ \mathbb{E} \left( \sum_{y} W_{n}^{n}(y|X_{1})1 \left\{ (X_{j}, y) \in A_{n} \right\} \right) \leq e^{-\gamma} \exp \left( \omega^{2}_{n}(\xi - 1) \right). \] (72)
We now analyze the first term on the right-hand side of \(61\).

\[
\mathbb{E}\left( \sum_y W^n_{Y|X}(y|X_1) \mathbf{1}\{(X_1, y) \notin \mathcal{A}^n_1\} \right) = P_{W_{Y|X}^n}(\ln \frac{W^n_{Y|X}(Y|X)}{P^n_0(y)} < \gamma) \\
= P_{W_{Y|X}^n}(\sum_{i=1}^n \ln \frac{W^n_{Y|X}(Y_i|X_i)}{P^n_0(Y_i)} < \gamma).
\]

(73)

If \(X_i = 0\), note that \(Y_i\) is distributed according to \(P_0\) and that \(\ln \frac{W^n_{Y|X}(Y_i|0)}{P^n_0(Y_i)} = 0\). Similarly, if \(X_i = 1\), then \(Y_i\) is distributed according to \(P_1\) and \(\ln \frac{W^n_{Y|X}(Y_i|1)}{P^n_0(Y_i)} = \ln \frac{P(Y_i)}{P^n_0(Y_i)}\). Consequently, although the sum in (73) contains \(n\) terms, only those for which \(X_i = 1\) contribute to it. Therefore, we introduce the random variable \(L \triangleq \sum_{i=1}^n 1\{X_i = 1\}\), so that

\[
P_{W_{Y|X}^n}\left(\sum_{i=1}^n \ln \frac{W^n_{Y|X}(Y_i|X_i)}{P^n_0(Y_i)} < \gamma\right) = \mathbb{E}_L\left[P_{W_{Y|X}^n}\left(\sum_{i=1}^n \ln \frac{W^n_{Y|X}(Y_i|X_i)}{P^n_0(Y_i)} < \gamma\right) \mid L\right] \\
= \mathbb{E}_L\left[P_{L}\left(\sum_{i=1}^L \ln \frac{P(Y_i)}{P^n_0(Y_i)} < \gamma\right) \mid L\right).
\]

(74)

Let \(\mu, \nu \in [0; 1]\) and set

\[
\gamma \triangleq (1 - \mu)(1 - \nu)\omega_n \sqrt{n} \mathbb{D}(P_1||P_0) \quad \text{and} \quad C^\mu_n \triangleq \{\ell \in \mathbb{N}^* : \ell > (1 - \mu)\omega_n \sqrt{n}\}.
\]

(75)

Then,

\[
\mathbb{E}_L\left[P\left(\sum_{i=1}^L \ln \frac{P(Y_i)}{P^n_0(Y_i)} < \gamma\right) \mid L\right] \leq \sum_{\ell \in C^\mu_n} P(L = \ell) P_{L}\left(\sum_{i=1}^\ell \ln \frac{P(Y_i)}{P^n_0(Y_i)} < \gamma\right) + P(L \notin C^\mu_n).
\]

(76)

Since \(\mathbb{E}(L) = \sum_{i=1}^n \mathbb{E}(1\{X_i = 1\}) = \omega_n \sqrt{n}\), we obtain with a Chernoff bound

\[
P(L \notin C^\mu_n) = P(L \leq (1 - \mu)\mathbb{E}(L)) \leq \exp\left(-\frac{\mu^2 \omega_n \sqrt{n}}{2}\right).
\]

(77)

For \(\ell \in C^\mu_n\), we have

\[
(1 - \mu)(1 - \nu)\omega_n \sqrt{n} - \ell < (1 - \mu)\ell - \ell = -\nu \ell
\]

so that

\[
P_{L}\left(\sum_{i=1}^\ell \ln \frac{P(Y_i)}{P^n_0(Y_i)} < \gamma\right) = P_{L}\left(\sum_{i=1}^\ell \ln \frac{P(Y_i)}{P^n_0(Y_i)} - \ell \mathbb{D}(P_1||P_0) < \gamma - \ell \mathbb{D}(P_1||P_0)\right) \\
\leq P_{L}\left(\sum_{i=1}^\ell \ln \frac{P(Y_i)}{P^n_0(Y_i)} - \ell \mathbb{D}(P_1||P_0) < -\nu \ell \mathbb{D}(P_1||P_0)\right) \\
\leq B \exp(-b\ell) \text{ for some constants } B, b > 0
\]

(79)

\[
\leq B \exp(-b\ell(1 - \mu)\omega_n \sqrt{n})\quad \text{,}
\]

(80)

where we have used the assumption that \(\ln \frac{P(Y)}{P^n_0(Y)}\) is sub-Gaussian. Combining, (74)-(80) with (73), we obtain

\[
\mathbb{E}\left( \sum_y W^n_{Y|X}(y|X_1) \mathbf{1}\{(X_1, y) \notin \mathcal{A}^n_1\} \right) \leq B \exp(-b\ell(1 - \mu)\omega_n \sqrt{n}) + \exp\left(-\frac{\mu^2 \omega_n \sqrt{n}}{2}\right).
\]

(81)

Using (72) and (81) with (61), we finally have

\[
\mathbb{E}(P_x) \leq B \exp(-b\ell(1 - \mu)\omega_n \sqrt{n}) + \exp\left(-\frac{\mu^2 \omega_n \sqrt{n}}{2}\right) + M e^{-\gamma} \exp(\omega^2_n(\xi - 1))
\]

(82)
Hence, if
\[ \ln M = (1 - \delta)(1 - \mu)(1 - \nu)D(P_1\|P_0)\omega_n\sqrt{n} \] with \( \delta \in [0; 1] \),
we obtain
\[
\mathbb{E}(P_e) \leq B \exp\left(-b\ell(1 - \mu)\omega_n\sqrt{n}\right) + \exp\left(-\frac{\mu^2\omega_n\sqrt{n}}{2}\right) + \exp\left(-\delta(1 - \mu)(1 - \nu)D(P_1\|P_0)\omega_n\sqrt{n}\right) \exp\left(\omega_n^2(\xi - 1)\right) \tag{84}
\]
For \( n \) large enough, \( \exp\left(\omega_n^2(\xi - 1)\right) \leq e \) so that
\[
\mathbb{E}(P_e) \leq \exp\left(-\rho_1\omega_n\sqrt{n}\right) \text{ for some appropriate choice of } \rho_1 > 0. \tag{85}
\]

h) Channel resolvability analysis: The channel resolvability analysis is similar to that developed in the proof of Theorem 2. In particular, one can show that for any \( \tau > 0 \)
\[
\mathbb{E}(V(Q_n^n, Q^n_{\alpha_n})) \leq \mathbb{P}_{W^n_{Z\|X}}\left(\sum_{i=1}^n \ln \frac{W_{Z\|X}(Z_i|X_i)}{Q_0(Z_i)} > \tau\right) + \frac{1}{2}\sqrt{\frac{2\tau}{M}}. \tag{86}
\]
Consequently, for \( \mu, \nu, \delta \) as fixed earlier, choosing
\[
\tau \triangleq (1 + \mu)(1 + \nu)D(Q_1\|Q_0)\omega_n\sqrt{n} \quad \text{and} \quad \ln M > (1 + \delta)(1 + \mu)(1 + \nu)D(Q_1\|Q_0)\omega_n\sqrt{n} \tag{87}
\]
and proceeding as in the proof of Theorem 2 ensures that
\[
\mathbb{E}(V(Q^n_n, Q^n_{\alpha_n})) \leq e^{-\rho_2\omega_n\sqrt{n}} \text{ with some appropriate } \rho_2 > 0. \tag{88}
\]

Choosing \( \mu, \nu, \delta \) so that \( \ln M \) satisfies both (83) and (87) and using Markov’s inequality with together with the bounds in (85) and (88), one may conclude that there exists a specific coding scheme with \( n \) large enough such that
\[
\mathbb{P}(W \neq \hat{W}) \leq 3e^{-\rho_1\omega_n\sqrt{n}} \quad \text{and} \quad V(Q^n_n, Q^n_{\alpha_n}) \leq 3e^{-\rho_2\omega_n\sqrt{n}}. \tag{89}
\]
One may then find appropriate constants to obtain the result promised in the statement of the theorem.

V. Extensions and Applications

A. Gaussian channels

Gaussian channels are of particular practical interest with the innocent symbol \( x_0 = 0 \). The architecture results developed in earlier section directly apply to this setting, upon checking that the condition of the theorems are satisfied. In particular, if \( P_i \sim \mathcal{N}(x_i, \sigma) \), one may check that \( \ln \frac{P_1(Y)}{P_0(Y)} \) for \( Y \sim P_1 \) is sub-Gaussian since
\[
\ln \frac{P_1(Y)}{P_0(Y)} = \frac{x_1}{\sigma^2}Y - \frac{x_1^2}{2\sigma^2}, \tag{90}
\]
which is a Gaussian random variable. In addition,
\[
\int \frac{P_1(y)^2}{P_0(y)} dy = e^{-\frac{\mu^2}{2}} < \infty. \tag{91}
\]
B. Covert and secret communication

The problem as formulated in Section II only requires communication to be undetectable, but does not prevent the warden from extracting information about the transmitted message. To address this, one could consider an additional semantic secrecy constraint of the form

$$\forall p_W \lim_{n \to \infty} I(W; Z^n) = 0.$$  \hspace{1cm} (92)

The problem is then similar to the effective secrecy introduced in [8] in a regime of undetectable communication.

The key-assisted architectures studied in Section III already satisfy this condition because the modulation as per (18) performs a one-time pad of the encoded message bits with the key bits $\hat{S}$. If one does not wish to use a key $S$, then the secrecy condition (92) may be obtained by using a code for the wiretap channel instead of a code for reliable communication. The result is then similar to Theorem 1 or Theorem 2 upon replacing the capacity $C$ by the secrecy capacity $C_s$.

In the keyless architecture, the secrecy condition may also be imposed by introducing a binning structure in the code and following the principle of achieving secrecy from resolvability. We omit the details here for brevity, but one may show that there exists a key-less covert communication scheme such that

$$\log_2 M = (1 - \alpha) (D(P_1 \parallel P_0) - D(Q_1 \parallel Q_0)),$$

$$\mathbb{P}(W \neq \hat{W}) \leq e^{-\rho_1 \sqrt{n}},$$  \hspace{1cm} (93)

$$\forall (\hat{Q}_n, Q_0^n) \leq e^{-\rho_2 \sqrt{n} + \omega_n \chi (Q_1 \parallel Q_0)},$$  \hspace{1cm} (94)

$$\forall p_W I(W; Z^n) \leq e^{-\rho_3 \sqrt{n}}.$$  \hspace{1cm} (95)

VI. ACKNOWLEDGEMENTS

The author would like to thank Boulat Bash for stimulating discussions and for introducing him to the problem of covert communications. This work was supported in part by the National Science Foundation with award CIF 1320298 and the Agence Nationale pour la Recherche under Grant ANR-13-BS03-0008.


APPENDIX A

PROOF OF LEMMA 2

Applying Lemma 1 repeatedly to the product distributions $Q^n_0$ and $Q^n$, we obtain (see also [17, Corollary 3.3.11])

$$\forall (Q^n, Q^n_0) \leq H(Q^n, Q^n_0) \leq \sqrt{n} H(Q, Q_0) \leq \sqrt{n} \chi(Q||Q_0),$$

(97)

which is non-trivial if $Q \ll Q_0$. Furthermore,

$$\chi(Q||Q_0) = \sqrt{\sum_{z \in Z} ((Q(z) - Q_0(z))^2}{Q_0(z)}} = \sqrt{\sum_{z \in Z} ((Q_1(z) + Q_0(z)(1 - \alpha_n) - Q_0(z))^2}{Q_0(z)}}$$

$$= \alpha_n \sqrt{\sum_{z \in Z} ((Q_1(z) - Q_0(z))^2}{Q_0(z)}}$$

$$= \alpha_n \chi(Q_1||Q_0).$$

(98)

Combining (97) and (98) yields the first result.

Next, note that

$$\mathbb{I}(P_X; W_{Z||X}) = (1 - \alpha_n) \mathbb{D}(Q_0||Q) + \alpha_n \mathbb{D}(Q_1||Q)$$

$$= (1 - \alpha_n) \mathbb{D}(Q_0||Q) + \alpha_n \mathbb{D}(Q_1||Q_0) + \alpha_n \sum_z Q_1(z) \ln \frac{Q_0(z)}{Q(z)}$$

$$= \alpha_n \mathbb{D}(Q_1||Q_0) - \mathbb{D}(Q||Q_0)$$

(99)

From Lemma 1 and immediate calculus we obtain

$$\mathbb{D}(Q||Q_0) \geq \mathbb{V}(Q, Q_0)^2 = \alpha_n^2 \mathbb{V}(Q_1, Q_0)^2.$$

(100)

In addition,

$$\mathbb{D}(Q||Q_0) = \sum_{z:Q_0(z)>0} (Q_0(z) + \alpha_n (Q_1(z) - Q_0(z))) \ln \left(1 + \alpha_n \frac{Q_1(z) - Q_0(z)}{Q_0(z)}\right)$$

$$\leq \sum_{z:Q_0(z)>0} (Q_0(z) + \alpha_n (Q_1(z) - Q_0(z))) \alpha_n \frac{Q_1(z) - Q_0(z)}{Q_0(z)}$$

$$= \alpha_n \sum_{z:Q_0(z)>0} (Q_1(z) - Q_0(z)) + \alpha_n^2 \sum_{z:Q_0(z)>0} \frac{(Q_1(z) - Q_0(z))^2}{Q_0(z)}$$

$$= \alpha_n^2 \chi(Q_1||Q_0)^2,$$

(101)

where (a) follows from the inequality $\ln(1 + x) < x$ for $x > 0$, and (b) follows because $\sum_{z:Q_0(z)>0} Q_1(z) = 1$ since $Q_1 \ll Q_0$. Combining (99), (100), and (101), we obtain the desired results.

APPENDIX B

PROOF OF LEMMA 3

The proof follows ideas from [21]. For $\gamma > 0$, define the set

$$A^n_\gamma \triangleq \left\{ x \in \{x_0, x_1\}^n : \ln \frac{1}{\Pi^n_{\alpha_n}(x)} < \gamma \right\}$$

(102)

and

$$\hat{P}^{(1)}(x) \triangleq \sum_{i=1}^K \frac{1}{K} \mathbf{1}\{x = \bar{x}_i\} \mathbf{1}\{\bar{x}_i \in A^n_\gamma\} \text{ and } \hat{P}^{(2)}(x) \triangleq \sum_{i=1}^K \frac{1}{K} \mathbf{1}\{x = \bar{x}_i\} \mathbf{1}\{\bar{x}_i \notin A^n_\gamma\}$$

(103)
so that \( P_{\tilde{X}} = \bar{P}^{(1)} + \bar{P}^{(2)} \). Taking the average over the random code generation, note that \( \mathbb{E}(P_{\tilde{X}}(\mathbf{x})) = \Pi_{\alpha_n, \epsilon}(\mathbf{x}) \), so that

\[
\mathbb{E}(\mathcal{V}(P_{\tilde{X}}, \Pi_{\alpha_n, \epsilon}(\mathbf{x}))) \leq \frac{1}{2} \sum_{\mathbf{x}} \mathbb{E}\left( \left| \bar{P}^{(1)}(\mathbf{x}) - \mathbb{E}\left( \bar{P}^{(1)}(\mathbf{x}) \right) \right| \right) + \frac{1}{2} \sum_{\mathbf{x}} \mathbb{E}\left( \left| \bar{P}^{(2)}(\mathbf{x}) - \mathbb{E}\left( \bar{P}^{(2)}(\mathbf{x}) \right) \right| \right).
\] (104)

Using Jensen’s inequality and the concavity of \( x \mapsto \sqrt{x} \), the first term on the right-hand side of (104) is bounded by \( \frac{1}{2} \sum_{\mathbf{x}} \sqrt{\text{Var}(\bar{P}^{(1)}(\mathbf{x}))} \) with

\[
\text{Var}(\bar{P}^{(1)}(\mathbf{x})) = \frac{1}{M} \text{Var}(1 \{ \mathbf{x} = \mathbf{X} \} 1 \{ \mathbf{X} \in A^n \})
= \frac{1}{\lambda_n K} \mathbb{E}_{\Pi_{\alpha_n}} \left( 1 \{ \mathbf{x} = \mathbf{X} \} 1 \{ \mathbf{X} \in A^n \} \right)
\leq \frac{1}{\lambda_n K} \Pi_{\alpha_n}(\mathbf{x})^2 e^{\gamma} 1 \{ \mathbf{x} \in A^n \}
\leq \frac{1}{\lambda_n K} \Pi_{\alpha_n}(\mathbf{x})^2 e^{\gamma}.
\] (105)

Therefore,

\[
\frac{1}{2} \sum_{\mathbf{x}} \mathbb{E}\left( \left| \bar{P}^{(1)}(\mathbf{x}) - \mathbb{E}(\bar{P}^{(1)}(\mathbf{x})) \right| \right) \leq \frac{1}{2} \sqrt{\frac{e^{\gamma}}{\lambda_n K}}.
\] (106)

Now, define \( \text{sup}(\mathbf{x}) \triangleq |i \in [1, n] : x_i = x_1| \). The second term on the right-hand side of (104) is bounded by \( \sum_{\mathbf{x}} \mathbb{E}(\bar{P}^{(2)}(\mathbf{x})) \) and

\[
\sum_{\mathbf{x}} \mathbb{E}(\bar{P}^{(2)}(\mathbf{x})) \leq \frac{1}{\lambda_n} \mathbb{E}_{\Pi_{\alpha_n}} \left( 1 \{ \mathbf{X} \notin A^n_{\gamma} \} \right)
= \frac{1}{\lambda_n} \mathbb{P}_{\Pi_{\alpha_n}} \left( \ln \frac{1}{\Pi_{\alpha_n}(\mathbf{X})} \geq \gamma \right)
= \frac{1}{\lambda_n} \mathbb{P}_{\Pi_{\alpha_n}} \left( \ln \frac{1}{\alpha_n^{\text{sup}(\mathbf{X})} (1 - \alpha_n)^{n - \text{sup}(\mathbf{X})}} \geq \gamma \right)
= \frac{1}{\lambda_n} \mathbb{P}_{\Pi_{\alpha_n}} \left( \text{sup}(\mathbf{X}) \ln \frac{1 - \alpha_n}{\alpha_n} - n \ln(1 - \alpha_n) \geq \gamma \right)
\geq \frac{\gamma + n \ln(1 - \alpha_n)}{\ln \frac{1 - \alpha_n}{\alpha_n}}.
\] (107)

**APPENDIX C**

**PROOF OF LEMMA 4**

For \( \gamma > 0 \), define the set\(^3\)

\[
\mathcal{B}^{\alpha_n}_{\gamma} \triangleq \left\{ (\mathbf{x}, \mathbf{z}) \in \mathcal{A}^n \times \mathcal{Z}^n : \ln \frac{W^w_{\mathbf{z} \mid \mathbf{x}}(\mathbf{z} | \mathbf{x})}{Q^{\alpha_n}_0(\mathbf{z})} < \gamma \right\}.
\] (108)

and

\[
\hat{Q}^{(1)}(\mathbf{z}) \triangleq \sum_{i=1}^{K} W^w_{\mathbf{z} \mid \mathbf{x}}(\mathbf{z} | \overline{x}_i) \frac{1}{K} 1 \{ (\overline{x}_i, \mathbf{z}) \in \mathcal{B}^{\alpha_n}_{\gamma} \}
\]
(109)

\[
\hat{Q}^{(2)}(\mathbf{z}) \triangleq \sum_{i=1}^{K} W^w_{\mathbf{z} \mid \mathbf{x}}(\mathbf{z} | \overline{x}_i) \frac{1}{K} 1 \{ (\overline{x}_i, \mathbf{z}) \notin \mathcal{B}^{\alpha_n}_{\gamma} \}
\] (1010)

\(^3\)A traditional typical set be similar to \( \mathcal{B}^{\alpha_n}_{\gamma} \) but with \( Q^n(\mathbf{z}) \) in place of \( Q^{\alpha_n}_0(\mathbf{z}) \).
so that $\hat{Q}^n = \hat{Q}^{(1)} + \hat{Q}^{(2)}$. Also note that $\mathbb{E}(\hat{Q}^n(z)) = Q^n_{\beta,n,e}(z)$. Hence,

$$\mathbb{E}(\mathbb{V}(\hat{Q}^n, Q^n_{\beta,n,e})) \leq \frac{1}{2} \sum_z \mathbb{E}\left(\left|\hat{Q}^{(1)}(z) - \mathbb{E}\left(\hat{Q}^{(1)}(z)\right)\right|\right) + \frac{1}{2} \sum_z \mathbb{E}\left(\left|\hat{Q}^{(2)}(z) - \mathbb{E}\left(\hat{Q}^{(2)}(z)\right)\right|\right). \tag{111}$$

The first term on the right-hand side of (111) is bounded by $\frac{1}{2} \sum_z \sqrt{\text{Var}(\hat{Q}^{(1)}(z))}$ with

$$\text{Var}(\hat{Q}^{(1)}(z)) = \sum_{i=1}^K \frac{1}{K^2} \mathbb{V}\left(W_{Z|X}^n(z|x)1\{(\hat{x}_i, z) \in \mathcal{B}_n^\alpha\}\right). \tag{112}$$

By Jensen’s inequality and the concavity of $x \mapsto \sqrt{x}$, we have

$$\sum_z \hat{Q}_0(z) \sqrt{\frac{\sum_x W_{Z|X}^n(z|x)\Pi_{n,\alpha}(x)}{Q_0(z)}} \leq \sqrt{\sum_z \sum_x W_{Z|X}^n(z|x)\Pi_{n,\alpha}(x) = 1} \tag{119}$$

so that

$$\frac{1}{2} \sum_z \mathbb{E}\left(\left|\hat{Q}^{(1)}(z) - \mathbb{E}\left(\hat{Q}^{(1)}(z)\right)\right|\right) \leq \frac{1}{2} \sum_z \frac{1}{\lambda_n K} \hat{Q}_0(z)^2 \frac{\sum_x W_{Z|X}^n(z|x)\Pi_{n,\alpha}(x)}{Q_0(z)} \tag{117}$$

$$\leq \frac{1}{2} \sqrt{\frac{e^\gamma}{\lambda_n K} \sum_z \hat{Q}_0(z)^2 \frac{\sum_x W_{Z|X}^n(z|x)\Pi_{n,\alpha}(x)}{Q_0(z)}} \tag{118}$$

By Jensen’s inequality and the concavity of $x \mapsto \sqrt{x}$, we have

$$\sum_z \hat{Q}_0(z) \sqrt{\frac{\sum_x W_{Z|X}^n(z|x)\Pi_{n,\alpha}(x)}{Q_0(z)}} \leq \sqrt{\sum_z \sum_x W_{Z|X}^n(z|x)\Pi_{n,\alpha}(x) = 1} \tag{119}$$

so that

$$\frac{1}{2} \sum_z \mathbb{E}\left(\left|\hat{Q}^{(1)}(z) - \mathbb{E}\left(\hat{Q}^{(1)}(z)\right)\right|\right) \leq \frac{1}{2} \sqrt{\frac{2\gamma}{\lambda_n K}}. \tag{120}$$

The second term on the right-hand side of (111) is upper bounded by $\sum_z \mathbb{E}(\hat{Q}^{(2)}(z))$ with

$$\sum_z \mathbb{E}(\hat{Q}^{(2)}(z)) = \sum_z \sum_{i=1}^K \sum_{\hat{x}_i} W_{Z|X}^n(z|\hat{x}_i)\hat{p}_{n,\alpha}(\hat{x}_i) \frac{1}{K} 1\{(\hat{x}_i, z) \notin \mathcal{B}_n^\alpha\}. \tag{121}$$

$$= \sum_z \sum_x W_{Z|X}^n(z|x)\hat{p}_{n,\alpha}(x) \frac{1}{K} 1\{(x, z) \notin \mathcal{B}_n^\alpha\}. \tag{122}$$

$$\leq \sum_z \sum_x W_{Z|X}^n(z|x)\hat{p}_{n,\alpha}(x) \frac{1}{\lambda_n} 1\{(x, z) \notin \mathcal{B}_n^\alpha\}. \tag{123}$$

$$= \frac{1}{\lambda_n} \mathbb{P}_{W_{Z|X}^n\Pi_{n,\alpha}}\left(\frac{\ln \frac{W_{Z|X}^n(Z|X)}{Q_0^n(Z)}}{Q_0^n(Z)} > \gamma\right). \tag{124}$$

$$= \frac{1}{\lambda_n} \mathbb{P}_{W_{Z|X}^n\Pi_{n,\alpha}}\left(\sum_{i=1}^n \ln \frac{W_{Z|X}^n(Z_i|X_i)}{Q_0^n(Z_i)} > \gamma\right). \tag{125}$$
REFERENCES

[1] B. 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, September 2013.
[2] B. Bash, S. Guha, D. Goeckel, and D. Towsley, “Quantum noise limited optical communication with low probability of detection,” in Proc. of IEEE International Symposium on Information Theory, Istanbul, Turkey, July 2013, pp. 1715–1719.
[3] A. D. Ker, “A capacity result for batch steganography,” IEEE Signal Processing Letters, vol. 14, no. 8, pp. 525–528, 2007.
[4] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, “Reliable deniable communication with channel uncertainty,” in Proc. of IEEE Information Theory Workshop, Hobart, Tasmania, November 2014, pp. 30–34.
[5] S. Lee, R. Baxley, J. McMahon, and R. Frazier, “Achieving positive rate with undetectable communication over MIMO rayleigh channels,” in Proc. of IEEE 8th Sensor Array and Multichannel Signal Processing Workshop, A Coruña, Spain, June 2014, pp. 257–260.
[6] S. Lee and R. Baxley, “Achieving positive rate with undetectable communication over AWGN and Rayleigh channels,” in Proc. of IEEE International Conference on Communications, Sydney, Australia, June 2014, pp. 780–785.
[7] B. Bash, D. Goeckel, and D. Towsley, “Lpd communication when the warden does not know when,” in Proc. IEEE International Symposium on Information Theory, Honolulu, Hawaii, July 2014, pp. 606–610.
[8] J. Hou and G. Kramer, “Effective secrecy: Reliability, confusion and stealth,” in Proc. of IEEE International Symposium on Information Theory, Honolulu, HI, July 2014, pp. 601–605.
[9] M. R. Bloch and J. N. Laneman, “Strong secrecy from channel resolvability,” IEEE Transactions on Information Theory, vol. 59, no. 12, pp. 8077–8098, December 2013.
[10] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, “Reliable, deniable and hidable communication: A quick survey,” in Proc. of IEEE Information Theory Workshop, 2014, pp. 227–231.
[11] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, “Hiding information in noise: Fundamental limits of covert wireless communication,” IEEE Communications Magazine, 2015, to appear.
[12] A. O. Hero, “Secure space-time communication,” IEEE Transactions on Information Theory, vol. 49, no. 12, pp. 3235–3249, December 2003.
[13] P. H. Che, M. Bakshi, and S. Jaggi, “Reliable deniable communication: Hiding messages in noise,” in Proc. of IEEE International Symposium on Information Theory, Istanbul, Turkey, July 2013, pp. 2945–2949.
[14] T. Han and S. Verdú, “Approximation theory of output statistics,” IEEE Transactions on Information Theory, vol. 39, no. 3, pp. 752–772, May 1993.
[15] T. S. Han, Information-Spectrum Methods in Information Theory. Springer, 2002.
[16] E. Lehmann and J. Romano, Testing Statistical Hypotheses. Springer, 2005.
[17] Reiss, Approximate Distributions of Order Statistics. Springer, 1989.
[18] S. Boucheron, G. Lugosi, and P. Massart, Concentration inequalities. Oxford University Press, 2013.
[19] G. Kramer, Topics in Multi-User Information Theory, ser. Foundations and Trends in Communications and Information Theory. NOW Publishers, 2008, vol. 4, no. 4-5.
[20] S. Verdú and T. S. Han, “A general formula for channel capacity,” IEEE Transactions on Information Theory, vol. 40, no. 4, pp. 1147–1157, July 1994.
[21] P. Cuff, “Distributed channel synthesis,” IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7071–7096, 2013.
[22] M. Bellare, S. Tessaro, and A. Vardy, “Semantic security for the wiretap channel,” in Advances in Cryptology - CRYPTO 2012, ser. Lecture Notes in Computer Science, R. Safavi-Naini and R. Canetti, Eds., vol. 7417. Springer Berlin Heidelberg, 2012, pp. 294–311, hard-copy.