Abstract—We consider a scenario in which $K$ transmitters attempt to communicate covert messages reliably to a legitimate receiver over a discrete memoryless multiple-access channel (MAC) while simultaneously escaping detection from an adversary who observes their communication through another discrete memoryless MAC. We assume that each transmitter may use a secret key that is shared only between itself and the legitimate receiver. We show that each of the $K$ transmitters can transmit on the order of $\sqrt{n}$ reliable and covert bits per $n$ channel uses, exceeding which, the warden will be able to detect the communication. We identify the optimal pre-constants of the scaling, which leads to a complete characterization of the covert capacity region of the $K$-user binary-input MAC. We show that, asymptotically, all sum-rate constraints are inactive unlike the traditional MAC capacity region. We also characterize the channel conditions that have to be satisfied for the transmitters to operate without a secret key.

Index Terms—Covert communication, low probability of detection, multi-user channels, multiple-access channels, physical-layer security, Shannon theory

I. INTRODUCTION

RECENTLY, there has been a renewed interest in the characterization of the information-theoretic limits of covert communication. Following the early work [2] identifying the existence of a square-root law similar to that of steganography [3], [4], several follow-up works have refined the characterization of the square-root law for point-to-point classical channels [5]–[8] and classical-quantum channels [9]–[11]; in particular, the covert capacity has been defined and precisely computed for Discrete Memoryless Channels (DMCs) and Additive White Gaussian Noise (AWGN) channels. Other extensions have included developing an understanding of when the square-root law does not apply, such as in the presence of channel uncertainty [12]–[15], [16] or timing uncertainty [17], [18]. There have also been several investigations of multi-user models [19], [20], [21], [22], timing channels [23], [24], artificial-noise assisted transmissions [25], and covert key generation [26], as well as code designs [27], [28], [29]. Finally, we note that related works have explored the concept of stealth [30], which is tightly tied to the notion of approximation of output statistics [31], [32] and may be viewed as low probability of interception, whereas covertness focuses on low probability of detection.

The main result developed in this paper is the characterization of the covert capacity region of the $K$-user binary-input MAC. The tools used are natural extensions of the techniques developed for point-to-point covert and stealth channels in [5], [6], [30] and for MAC resolvability [33]–[35], but the converse proof requires special care beyond the approach used in [6]. We extend our previous work [1] by analyzing $K$-user MACs for any $K \geq 2$ and characterizing the optimal key throughput required for covert communication. We show that, asymptotically, there exist no sum-rate constraints unlike the traditional MAC rate region; intuitively, this happens because covertness is such a stringent constraint that the covert users never transmit enough bits to saturate the capacity of the channel. The system behaves as if a covert communication budget were merely allocated to the different users. A similar behavior was observed [36, Theorem 6] in the calculation of the channel capacity per unit cost of a two-user MAC when both users consist of a free input symbol.

The remainder of the paper is organized as follows. In Section II, we set the notation used in the paper, and in Section III, we formally introduce our channel model and define the covert capacity region. In Section IV, we develop a preliminary result that captures the essence of our approach to covertness and extends [6, Lemma 1]. We establish the covert capacity region of the $K$-user binary-input MAC in Section V and conclude our work with a brief discussion of extensions and open problems in Section VI. The proofs of all lemmas are relegated to the appendix.

II. NOTATION

We denote random variables and their realizations in upper and lower case, respectively. All sequences in boldface are $n$-length sequences, where $n \in \mathbb{N}^*$, unless specified otherwise. We define the weight of a sequence as the number of non-zero symbols in that sequence. Throughout the paper, log and exp are understood to be the base $e$; the results can be interpreted in bits by converting log to the base 2. Adhering to standard information-theoretic notation, $H(X)$ and $I(X; Y)$ represent the entropy of $X$ and the mutual information between $X$ and $Y$, respectively. For $x \in [0, 1]$, let $\mathbb{H}_b(x)$ denote the binary
III. CHANNEL MODEL

We define the set \( K \triangleq [1, K] \), where \( K \in \mathbb{N}^* \) and \( K \geq 2 \). We analyze the channel model illustrated in Figure 1, in which \( K \) transmitters simultaneously communicate with a legitimate receiver over a discrete memoryless MAC \((X[K], W_{X|X[K]}, \mathcal{Y})\) in the presence of a warden monitoring the communication over another discrete memoryless MAC \((X[K], W_{Z|X[K]}, \mathcal{Z})\). As both channels are memoryless, we denote the transition probabilities corresponding to \( n \) uses of the channel by \( W_{Y|X[K]}^{\text{un}} \triangleq \prod_{i=1}^{n} W_{Y|X[K]} \) and \( W_{Z|X[K]}^{\text{un}} \triangleq \prod_{i=1}^{n} W_{Z|X[K]} \). In addition, we assume for simplicity of exposition that each user \( k \in K \) uses the same binary input alphabet \( \mathcal{X}_k \triangleq \mathcal{X} \triangleq \{0, 1\} \) and that the output alphabets \( \mathcal{Y} \) and \( \mathcal{Z} \) are finite. We let \( 0 \in \mathcal{X} \) be the innocent symbol corresponding to the channel input when no communication takes place. We assume that all terminals are synchronized and possess complete knowledge of the coding scheme used.

The user indexed by \( k \in K \) encodes a uniformly-distributed message \( W_k \in [1, M_k] \) and a uniformly-distributed secret key \( S_k \in [1, L_k] \), which is shared only with the receiver, into a codeword \( X_k(W_k, S_k) \in \mathcal{X}^n \) of length \( n \). We denote the collection of the \( K \) codewords \((X_k(W_k, S_k))_{k \in K}\) by \( X_K(W[K], S[K]) \). When the context is clear, we drop the message and key indices, \( W_k \) and \( S_k \), and denote \( X_k(W_k, S_k) \) by \( X_k \) instead for conciseness. It is convenient to think about the \( K \) inputs to the channel over \( n \) uses as a matrix \( X[K] \) of size \( K \times n \) obtained by vertically stacking the \( K \) codewords, each of which is a row vector. The inputs corresponding to all users indexed by the elements of a non-empty set \( U \subseteq K \) is a sub-matrix of \( X[K] \) obtained by selecting the rows whose indices belong to \( U \) and is denoted by \( X[U] \). The \( K \) users then transmit codewords \( X[K] \) over the channel in \( n \) channel uses. At the end of transmission, the receiver observes \( Y \) while the warden observes \( Z \), both of which are of length \( n \).

We introduce a \( K \)-length row vector \( X_U \equiv (X_1, X_2, \ldots, X_K), U \subseteq K \), with entry \( X_k \) if \( k \in U \) and \( X_k = 0 \) otherwise. With our assumption that all channel inputs are binary, we represent every column of the matrix \( X[K] \) by a vector \((X_U)^T\), where the set \( U \) consists of the indices of all users transmitting symbol 1 in this column. We denote the \( k \)-th component of \( X_U \) by \( X_{U,k} \). In accordance with the notation introduced in the previous paragraph, \( X_U[T] \) represents a row vector of length \( |T| \) that contains the entries \( \{X_{U,k}\}_{k \in T} \). Note the difference between \( X[U] \) and \( X_U \); the former is a \( |U| \)-length vector \( X_U \) whereas the latter is a \( K \)-length vector with 1’s in indices that belong to the set \( U \). For conciseness, we define

\[
P_U(y) \triangleq W_{Y|X[K]}(y|x_U), \quad Q_U(z) \triangleq W_{Z|X[K]}(z|x_U),
\]

which represent the one-shot output distributions at the legitimate receiver and the warden, respectively, when only the transmitters in \( U \subseteq K \) transmit symbol 1, while the transmitters in \( U^c \) transmit a 0. When \( U = \emptyset \), which occurs when all users transmit the innocent symbol 0, we write \( P_0 \) and \( Q_0 \). We assume that \( Q_U \ll Q_0 \) for all non-empty sets \( U \subseteq K \) and that \( Q_0 \) cannot be written as a convex combination of the form \( Q_k(z) = \sum_{\ell \subseteq K} \left( \prod_{k \in K} \mu_k \left( \prod_{k \in U} (1 - \mu_k) \right) \right) z_{k} \) for some \( \mu_k \in [0, 1] \). In the former case, covert communication involving all \( K \) users is impossible; in the latter case, covert communication would directly follow from known channel resolvability results [33]-[35] and would be possible at a non-zero rate. We also assume that there does not exist \( (\rho_k)_{k \in K} \in [0, 1]^K \) with \( \sum_{k \in K} \rho_k = 1 \) such that \( \sum_{k \in K} \rho_k Q_k(z) = Q_0(z) \) for all \( z \in \mathcal{Z} \). As we shall see later in Section IV, the square root law of covert communication can be circumvented if such a \((\rho_k)_{k \in K}\) exists.

Upon observing \( Y \), the legitimate receiver estimates the message vector \( \hat{W}[K] \). We measure reliability at the receiver with the average probability of error \( P_e \triangleq \mathbb{P}(\hat{W}[K] \neq W[K]) \). Upon observing \( Z \), the warden attempts to detect whether all \( K \) users transmitted covert messages (Hypothesis \( H_1 \)) or not (Hypothesis \( H_0 \)) by performing a hypothesis test on \( Z \). We denote the Type I (rejecting \( H_0 \) when true) and Type II (accepting \( H_0 \) when false) error probabilities by \( \alpha \) and \( \beta \), respectively. The warden can achieve any pair \((\alpha, \beta)\) such that \( \alpha + \beta = 1 \) by ignoring his observation \( Z \) and basing his decision on the result of a coin toss. We define the distribution induced at the warden when communication takes place by

\[
\tilde{Q}^{\text{un}}(z) \triangleq \frac{1}{\prod_{k \in K} M_k L_k} \sum_{m[K]} \sum_{\ell[K]} W_{Z|X[K]}(z|x_K) (m[K], \ell[K]).
\]  

(2)

We measure coverts in terms of the KL divergence \( \mathcal{D}(\tilde{Q}^{\text{un}} || Q_0^{\text{un}}) \), where \( Q_0^{\text{un}} \) is the distribution observed by the
warden when none of the K users transmits any covert information. We know from [38] that any test conducted by the warden on Z satisfies \( a + \beta \geq 1 - \sqrt{D(\Omega^n \parallel Q_{\theta}^n)} \). Using Pinsker's inequality [37], we write \( a + \beta \geq 1 - \sqrt{D(\Omega^n \parallel Q_{\theta}^n)} \). The primary objective of our covert communication scheme is to guarantee that \( D(\Omega^n \parallel Q_{\theta}^n) \) is negligible so that any statistical test used by the warden on Z is futile. Note that we only consider communication schemes for which \( \log M_k \) grows to infinity, for \( k \in K \), as \( n \) grows to infinity.

**Definition 1.** The tuple \( [K] \in \mathbb{R}^+ \) is an achievable reliable and covert throughput tuple if there exists a sequence of codes as defined above with increasing blocklength \( n \) such that for every \( k \in K \),

\[
\lim \inf_{n \to \infty} \frac{\log M_k}{\sqrt{nD(\Omega^n \parallel Q_{\theta}^n)}} \geq r_k,
\]

and

\[
\lim_{n \to \infty} P^a_n = 0, \quad \lim_{n \to \infty} D(\Omega^n \parallel Q_{\theta}^n) = 0.
\]

The covert capacity region of the K-user MAC consists of the closure of the set of all achievable throughput tuples \( r[K] \). Also, we define the tuple \( s[K] \in \mathbb{R}^+ \) as an achievable key throughput tuple associated with the achievable reliable and covert throughput tuple \( r[K] \), if there exist a sequence of codes satisfying (3) and (4) and if for all \( k \in K \),

\[
s_k \geq \lim \sup_{n \to \infty} \frac{\log L_k}{\sqrt{nD(\Omega^n \parallel Q_{\theta}^n)}}.
\]

Note that in (3), we normalize the number of bits \( \log M_k \) by \( \sqrt{nD(\Omega^n \parallel Q_{\theta}^n)} \) instead of \( n \) as traditionally done in information-theoretic problems. The normalization by \( \sqrt{n} \) is essential to reflect the fact that covert communication corresponds to a zero-rate regime, in which the number of bits scales sub-linearly with the number of channel uses. The normalization by \( \sqrt{D(\Omega^n \parallel Q_{\theta}^n)} \) is also crucial to reflect the fact that \( D(\Omega^n \parallel Q_{\theta}^n) \) influences \( \{\log M_k\}_{k \in K} \). While the normalization might seem somewhat ad-hoc, it is justified *a posteriori* in Section V when we prove that \( \log M_k / \sqrt{nD(\Omega^n \parallel Q_{\theta}^n)} \) is independent of \( n \) in the limit of large blocklength. Said differently, \( \log M_k / \sqrt{nD(\Omega^n \parallel Q_{\theta}^n)} \) plays the role of the usual “rate” in that it asymptotically does not depend on the blocklength \( n \) and already integrates the scaling. To avoid confusion, we refer to \( r_k \) as throughput instead of rate.

### IV. Preliminaries

Following the approach proposed in [6], we introduce a covert communication process, which is an independent and identically distributed (i.i.d.) process indistinguishable from the innocent distribution \( Q_{\theta}^n \) in the limit. The rationale for introducing this process is to precisely quantify the fraction of channel uses in which the users can transmit symbol 1 while simultaneously avoiding detection by the warden, without introducing the coding aspect of the problem yet.

For \( n \in \mathbb{N}^+ \), let \( a_n \in (0, 1) \). Let \( \rho \triangleq \{\rho_k\}_{k \in K} \in [0, 1]^K \) such that \( 3 \sum_{k \in K} \rho_k = 1 \). We define the input distributions \( \{X_k\}_{k \in K} \) on \( X \) as

\[
\Pi_{X_k}(1) = 1 - \Pi_{X_k}(0) = \rho_k a_n.
\]

The output distributions at the legitimate receiver and the warden when the input distribution of each user \( k \) is \( \Pi_{X_k} \) are defined, respectively, as

\[
P_{a_n}(y) = \sum_{x[K]} W_{X[K]}(y|X[K]) \left( \prod_{k \in K} \Pi_{X_k}(x_k) \right),
\]

\[
Q_{a_n}(z) = \sum_{x[K]} W_{Z[K]}(z|x[K]) \left( \prod_{k \in K} \Pi_{X_k}(x_k) \right).
\]

The \( n \)-fold product distributions corresponding to (6), (7), and (8) are

\[
\Pi_{X_k}^n = \prod_{j=1}^n \Pi_{X_k}, \quad P_{a_n}^n = \prod_{j=1}^n P_{a_n}, \quad Q_{a_n}^n = \prod_{j=1}^n Q_{a_n}.
\]

For a set \( T \subseteq K \), we define

\[
G_T(z) = \sum_{U \subseteq T} (-1)^{|T| - |U|} Q_U(z).
\]

Then, using Lemma 5 in Appendix A, we write

\[
Q_{a_n}(z) = Q_{\theta}(z) + \sum_{T \subseteq K : T \neq \emptyset} \left( \prod_{k \in T} \rho_k a_n \right) G_T(z).
\]

Note that since \( Q_T \ll Q_{\theta} \) for all non-empty sets \( T \subseteq K \), it is also true that \( Q_{a_n} \ll Q_{\theta} \). Furthermore, we define

\[
\zeta_n(z) \triangleq \frac{Q_{a_n}(z) - Q_{\theta}(z)}{a_n}, \quad \chi_n(\rho) \triangleq \sum_{\zeta(\cdot)} \zeta_n^2(\cdot) Q_{\theta}(\cdot),
\]

\[
\zeta(z) \triangleq \sum_{k \in K} \rho_k (Q_k(z) - Q_{\theta}(z)), \quad \chi(\rho) \triangleq \sum_{\zeta(\cdot)} \zeta^2(\cdot) Q_{\theta}(\cdot).
\]

In the following lemma, we bound the KL divergence between \( Q_{a_n} \) and \( Q_{\theta} \). Later, we use the results of this lemma to show that for specific choices of \( a_n \), the stochastic process \( Q_{a_n}^n \) is indistinguishable from the innocent distribution \( Q_{\theta}^n \) in the limit.

**Lemma 1.** Let the sequence \( \{a_n\}_{n \geq 1} \) be such that \( \lim_{n \to \infty} a_n = 0 \). Then, for \( n \in \mathbb{N}^+ \) large enough,

\[
\frac{a_n^2}{2} (1 + a_n) \chi_n(\rho) \geq \mathbb{D}(Q_{a_n} \parallel Q_{\theta}) \geq \frac{a_n^2}{2} (1 - a_n) \chi_n(\rho).
\]

In addition, for all \( z \in Z \), \( \lim_{n \to \infty} \zeta_n(z) = \zeta(z) \) and \( \lim_{n \to \infty} \chi_n(\rho) = \chi(\rho) \). Finally, for random variables \( (X[T], \mathcal{X}[T]) \), \( Z \in \mathbb{X}_{[T]} \times \mathbb{Z} \) for some non-empty set \( T \subseteq K \) with joint distribution \( W_{Z[X[T]]} \left( \prod_{k \in T} \Pi_{X_k} \right) \), we have

\[
\mathbb{D}(X[T] ; Z) = \sum_{k \in T} \rho_k a_n \mathbb{D}(Q_k \parallel Q_{\theta}) + O\left(a_n^2\right).
\]
The proof of Lemma 1 is provided in Appendix B. Assume that each transmitter \( k \in K \) generates a sequence of length \( n \) using the process \( \Pi_{X_k}^{\alpha n} \). The weight of these sequences is \( \rho_k n a_n \) on average. To be indistinguishable from the innocent distribution in the limit, the covert process \( Q_{a_n}^{\alpha n} \) has to satisfy
\[
\lim_{n \to \infty} \mathbb{D}(Q_{a_n}^{\alpha n} \mid Q_0^{\alpha n}) = \lim_{n \to \infty} n \mathbb{D}(Q_{a_n} \mid Q_0) = 0. \tag{16}
\]
Our assumptions in Section III ensure that \( \chi(\rho) \) is non-zero. Consequently, from the results of Lemma 1 and (16), we conclude that if we choose the sequence \( \{a_n\}_{n \in \mathbb{N}} \) such that \( \lim_{n \to \infty} n a_n^2 = 0 \), our covert process \( Q_{a_n}^{\alpha n} \) is indistinguishable from \( Q_0^{\alpha n} \) in the limit. Consequently, we will construct a coding scheme that emulates the covert process \( Q_{a_n}^{\alpha n} \) instead of \( Q_0^{\alpha n} \). The prime benefit of using \( Q_{a_n}^{\alpha n} \) instead of \( Q_0^{\alpha n} \) is that \( Q_{a_n}^{\alpha n} \) allows us to convey covert information through the use of 1 symbols. In particular, it is possible to choose \( \{a_n\}_{n \in \mathbb{N}} \) such that \( \lim_{n \to \infty} n a_n = \infty \) so that the number of information bits grows with \( n \).

The “square root law” of covert communication follows from the constraint \( \lim_{n \to \infty} n a_n^2 = 0 \), which forces the scaling of \( n a_n \) to be arbitrarily close to but not exceed \( \sqrt{n} \). If \( \chi(\rho) = 0 \) for some \( \rho \), one would need to push the approximation of \( \mathbb{D}(Q_{a_n}^{\alpha n} \mid Q_0^{\alpha n}) \) at least to the order \( a_n^2 \) in Lemma 1. In turn, we would only need to choose a sequence such that \( \lim_{n \to \infty} n a_n^3 = 0 \), effectively allowing the increase of the scaling of \( n a_n \) to be arbitrarily close to but not exceed \( n^{2/3} \) and beating the square root law. The assumption that \( \chi(\rho) > 0 \) made in Section III therefore excludes the (rare) situations in which the square root law can be beaten.

V. MAIN RESULT

We characterize the covert capacity region of a \( K \)-user binary-input MAC in Theorem 1, with the achievability proof in Section V-B and the converse proof in Section V-C. The proofs adapt channel resolvability and converse techniques used in [6] for point-to-point channels to the MACs. The achievability proof is an extension of [6], and we provide details in the appendix; the converse proof presents more challenges and is fully detailed.

A. Covert Capacity Region of the \( K \)-User Binary-Input MAC

\textbf{Theorem 1.} For \( \rho \triangleq \{\rho_k\}_{k \in K} \in [0,1]^K \) such that \( \sum_{k \in K} \rho_k = 1 \), define
\[
\chi(\rho) \triangleq \sum_z \left( \sum_{k \in K} \rho_k (Q_k(z) - \rho \chi(z)) \right)^2. \tag{17}
\]
For the \( K \)-user binary-input MAC described in Section III, the covert capacity region is (18) shown at the bottom of this page. In addition, for any achievable reliable and covert throughput tuple \( r[K] \) on the boundary of the covert capacity region characterized by \( \rho \), the set of achievable key throughput tuples is (19) shown at the bottom of this page.

Note that \( \chi(\rho) \) in (17) is positive under the assumption made in Section III, so that the bounds in (18) and (19) are well defined and finite. A few remarks are now in order.

- Our characterization of the covert capacity region only involves constraints on individual user’s throughputs; there are no active constraints on the sum throughput. However, the individual throughputs are not identical to those of the single-user case [6], as there exists a non-trivial interplay among the \( \rho_k \)’s, for \( k \in K \), through \( \chi(\rho) \) in (18).
- User \( k \in K \) can achieve its maximum covert and reliable throughput without a key only if
\[
\mathbb{D}(P_k \parallel P_0) \geq \mathbb{D}(Q_k \parallel Q_0), \tag{20}
\]
is satisfied; that is, no secret key is required for user \( k \) if the channel from user \( k \) to the receiver is better than the channel to the warden when all other users are silent.
- If the MAC is symmetric, in the sense that \( \forall z \in Z \) and \( \forall k \in K \), \( Q_k(z) = Q(z) \), then \( \sum_{k \in K} \rho_k (Q_k(z) - Q_0(z)) = Q(z) - Q_0(z) \), so that \( \chi(\rho) \) is independent of \( \rho \) and time sharing is optimal.

Figure 2 illustrates the covert capacity region for a 2-user MAC with randomly generated channel matrices, \( W_{X_1X_2} \) and \( W_{Z|X_1X_2} \), that satisfy (20) for \( k \in \{1,2\} \) and the absolute continuity requirements described in Section III.
for $\mathcal{K} = \{1, 2\}$. The thick solid curve denotes the boundary of the covert capacity region. All points on this boundary can be achieved by varying the values of $(\rho_1, \rho_2)$. For $\rho = \rho^* \triangleq (\rho_1^*, \rho_2^*)$, the achievable covert throughput region is highlighted in Figure 2, where the square marker represents the maximum achievable covert throughput pair $\left(\frac{2}{\sqrt{\mathcal{P}}} \rho_1^* \mathbb{D}(P_1 || P_0), \frac{2}{\sqrt{\mathcal{P}}} \rho_2^* \mathbb{D}(P_2 || P_0)\right)$, while the triangular marker represents the pair $\left(\frac{2}{\sqrt{\mathcal{P}}} \rho_1^* \mathbb{D}(Q_1 || Q_0), \frac{2}{\sqrt{\mathcal{P}}} \rho_2^* \mathbb{D}(Q_2 || Q_0)\right)$. A non-empty intersection of the region to the right-top of the triangular marker and the region to the bottom-left of the square marker implies the existence of keyless covert communication schemes. If the regions do not intersect, a secret key is required to communicate covertly. Note that the achievable region is still the region spanning from $(0,0)$ to the square marker as highlighted in Figure 2. Also note that, for a symmetric 2-user MAC, the boundary of the covert capacity region is a straight line, and time sharing is optimal.

B. Achievability Proof

We consider a communication scheme in which every user $k$ employs $L_k$ sub-codebooks, each consisting of $M_k$ codewords. The value of the key $S_k \in [1, L_k]$ chooses the sub-codebook that user $k$ uses to encode its message $W_k \in [1, M_k]$. The decoder, which possesses complete knowledge of the keys $S[\mathcal{K}]$, attempts to decode the messages sent by the $K$ transmitters. The idea underlying the scheme is to use channel resolvability techniques to ensure that the total number of codewords is sufficiently large to keep the warden confused, while simultaneously ensuring that each sub-codebook is small enough for the receiver to reliably decode the messages.

**Proposition 1.** Let $\rho \triangleq (\rho_k)_{k \in \mathcal{K}} \in [0, 1]^K$ with $\sum_{k \in \mathcal{K}} \rho_k = 1$. Let $(a_n)_{n \in \mathbb{N}}$ be such that $a_n \in (0, 1)$, $\lim_{n \to \infty} n a_n = \infty$, and $\lim_{n \to \infty} n a_n^2 = 0$. For the channel model described in Section III, for an arbitrary $\mu \in (0, 1)$, there exist covert communication schemes such that for all $k \in \mathcal{K}$,

$$r_k = (1 - \mu) \sqrt{\mathcal{P}} \rho_k \mathbb{D}(P_k || P_0),$$

$$s_k = \sqrt{\mathcal{P}} \rho_k \left( (1 + \mu) \mathbb{D}(Q_k || Q_0) - (1 - \mu) \mathbb{D}(P_k || P_0) \right)^+, \quad (22)$$

$$\lim_{n \to \infty} P^n_e = 0,$$

$$\lim_{n \to \infty} \mathbb{D}(\tilde{Q}^n || Q^m_0) = 0.$$  

**Proof:** To prove Proposition 1, we rely on random coding arguments for channel reliability and channel resolvability. However, the use of low-weight codewords in our communication scheme requires that we handle concentration inequalities carefully. Since basic concentration inequalities do not apply in the low-weight regime [6], we use Bernstein’s inequality to establish our random coding arguments. The proof follows otherwise along the lines of [6, Theorem 2].}

\begin{itemize}
  \item \textbf{Random codebook generation:} At each transmitter $k \in \mathcal{K}$, generate $M_k L_k$ codewords $x_k (m_k, \ell_k) \in \mathcal{X}^m$, where $(m_k, \ell_k) \in [1, M_k] \times [1, L_k]$, independently at random according to the distribution $\Pi_{x_k}^{m, L_k}$. For a set $T \subseteq \mathcal{K}$, define

$$W^n_{Y_{T[X_T]}(y|x[T])} \triangleq \sum_{x[T^c]} W^n_{Y_{[X[T]}(y|x[X]))} \left( \prod_{k \in T^c} \Pi_{x_k}^{m, L_k} (x_k) \right).$$

\end{itemize}

Note that $W^n_{Y_{T[X_T]}(y|x[T])}$ is a product distribution since each user $k \in \mathcal{K}$ generates its codeword according to an $n$-fold product distribution $\Pi_{x_k}^{m, L_k}$. Also, note that if $T = \emptyset$, $W^n_{Y_{X[T]}(y|x[T])} = B^\alpha_{\mathcal{K}}$. Define the set $A^n_\mathcal{T} \triangleq \bigcap_{T \in \mathcal{T}} A^n_T$ with $A^n_T$ as shown in (26) shown at the top of the next page, where, for every non-empty set $T \subseteq \mathcal{K}$, $\gamma_T \triangleq (1 - \mu)n[I(X[T]; Y[I[X[T^c]])$ for an arbitrary $\mu \in (0, 1)$. Encoder $k \in \mathcal{K}$ uses the key $S_k = \ell_k$ to map the message $W_k = m_k$ onto the codeword $x_k (m_k, \ell_k)$. The codewords are then transmitted through the memoryless MAC to the legitimate receiver. The decoder, who observes $y$ and has complete knowledge of the keys $\ell[\mathcal{K}]$, operates as follows.

- If there exists a unique $m[\mathcal{K}] \in \times_{k \in \mathcal{K}} [1, M_k]$ such that $(x_K (m_k, \ell_k), Y[I[X[T^c]])$, output $W[\mathcal{K}] = m[\mathcal{K}]$.
- Else, declare a decoding error.

\begin{itemize}
  \item \textbf{Channel resolvability analysis:} The decoding error probability $P^n_e$ averaged over all random codebooks satisfies the following.

\textbf{Lemma 2.} For any $\mu \in (0, 1)$, an $n$ large enough, and

$$\log M_k = (1 - \mu) \rho_k n a_n \mathbb{D}(P_k || P_0),$$

for every $k \in \mathcal{K}$, the probability of decoding error averaged over all random codebooks satisfies

$$\mathbb{E}(P^n_e) \leq \exp(-\xi n a_n),$$

for an appropriate $\xi > 0$.

The proof of Lemma 2 is provided in Appendix D.

\end{itemize}

\begin{itemize}
  \item \textbf{Channel resolvability analysis:} In the following lemma, we show that the KL divergence between the induced distribution and the covert stochastic process averaged over all random codebooks vanishes in the limit.

\textbf{Lemma 3.} For any $\mu \in (0, 1)$, an $n$ large enough, and

$$\log M_k = (1 + \mu) \rho_k n a_n \mathbb{D}(Q_k || Q_0),$$

for every $k \in \mathcal{K}$, the KL divergence between $\tilde{Q}^n$ and $Q^m_0$ averaged over all random codebooks satisfies

$$\mathbb{E}(\mathbb{D}(\tilde{Q}^n || Q^m_0)) \leq \exp(-\xi n a_n),$$

for an appropriate $\xi > 0$.

The proof of Lemma 3 is provided in Appendix E.
Using (14), (32), (34), and our choice of ARUMUGAM AND BLOCH: COVERT COMMUNICATION OVER A
that the covert capacity region contains the region defined
inequality, we obtain

\[ fies (23) and (24). Combining (14) and (34) yields \]

\[ \xi \]

\[ D \]

\[ \alpha \]

\[ \rho \]

\[ n \]

\[ \log \]

\[ M_k \]

\[ (38) as shown at the bottom of this page. In addition, any
achievable covert throughput tuple \([ x ]\) that is characterized by a specific \( \rho \) and lies on the boundary of the region defined in (39) is associated with an achievable key throughput tuple

\[ \left\{ \frac{2}{\sqrt{x}} \rho_k \left( D(Q_k || Q_\emptyset) - D(P_k || P_\emptyset) \right) \right\} \] _k \in \mathbb{K}. \]

C. Converse Proof

Proposition 2. For the channel model described in Section III, consider a sequence of covert communication schemes with increasing blocklength \( n \in \mathbb{N}^* \) characterized by \( \epsilon_n \triangleq P^n_e \) and \( \delta_n \triangleq \mathbb{D}(\hat{Q}^n || Q^{\emptyset}_n) \) such that \( \lim_{n \to \infty} \epsilon_n = 0 \) and \( \lim_{n \to \infty} \delta_n = 0 \). Then, there exists a vector \( \rho \triangleq \{ \rho_k \}_{k \in \mathbb{K}} \in [0, 1]^{k} \) with \( \sum_{k \in \mathbb{K}} \rho_k = 1 \) and an infinite subset \( \mathcal{N} \subseteq \mathbb{N}^* \), such that for all \( k \in \mathbb{K}, \)

\[ \liminf_{n \to \infty} \frac{\log M_k}{\sqrt{n} \rho_n} \leq \frac{2}{\sqrt{\chi(\rho)}} \rho_k \mathbb{D}(Q_k || Q_\emptyset). \]

(40)

For a sequence of codes that achieves the right hand side of (40) for all \( k \in \mathbb{K}, \) we have

\[ \limsup_{n \to \infty} \frac{\log M_k L_k}{\sqrt{n} \rho_n} \geq \frac{2}{\sqrt{\chi(\rho)}} \rho_k \mathbb{D}(Q_k || Q_\emptyset). \]

(41)

for all \( k \in \mathbb{K}. \)

Proof: Consider a sequence of covert communication schemes with increasing blocklength \( n \in \mathbb{N}^* \) characterized by \( \epsilon_n \triangleq P^n_e \) and \( \delta_n \triangleq \mathbb{D}(\hat{Q}^n || Q^{\emptyset}_n) \), and log \( M_k \) takes the maximum value such that \( \lim_{n \to \infty} \log M_k = \infty \) for all \( k \in \mathbb{K}. \) Each user \( k \) transmits an \( n \)-length codeword \( X_k = (X_{k1}, X_{k2}, \ldots, X_{kn}) \in \mathcal{X}^n \), where \( n \in \mathbb{N}^* \), to the receiver. For \( j \in [1, n] \), we denote the distribution of each symbol \( X_{kj} \) on \( \mathcal{X} \) by \( \pi_{X_{kj}} \), where

\[ \pi_{X_{kj}}(x) \triangleq \frac{\sum_{m_{kj} = 1}^{M_k} \pi_{kj}^{(n)} = 1} {M_k L_k} \mathbb{I} \{ X_{kj}(m_k, \ell_k) = x \}. \]

(42)

We define \( \Pi_{X_{kj}}(1) = 1 - \Pi_{X_{kj}}(0) \triangleq \mu_{kj}^{(n)}. \) Note that \( \mu_{kj}^{(n)} \) depends on \( n \), the transmitter index \( k \), and the symbol position \( j \). For any \( n \in \mathbb{N}^* \), we define a permutation \( \pi_{kj}^{(n)} \) of \([1, n]\) to define a new code such that

\[ (k^*, 1) = \arg \max_{(k, j) \in \mathbb{K} \times [1, n]} \mu_{kj}^{(n)}. \]

(43)
Since the channel is memoryless, the performance of the new code that satisfies (43) matches that of the original code. Hence, without loss of generality, we only study the sequence of codes for which (43) holds for every $n \in \mathbb{N}^*$. Note that the sequence $\{\mu_k^{(n)}\}_{k \in \mathcal{K}}$ belongs to $[0, 1]^\mathcal{K}$ which is a closed and bounded set. Hence, we can extract a convergent subsequence $\{\mu_k^{(n)}\}_{k \in \mathcal{K}^*}$, where $\mathcal{N}^* \subseteq \mathbb{N}^*$ is an infinite set, with limit $\mu_k^{(n)}$. Let us now assume that the sequence $\{\mu_k^{(n)}\}_{k \in \mathcal{K}}$ belongs to $[0, 1]^\mathcal{K}$ is not an all-zero sequence.

For $j \in \llbracket 1, n \rrbracket$, we denote the $K$-length vector $\{x_{kj}\}_{k \in \mathcal{K}}$ by $x_{(j)}[\mathcal{K}]$. The warden makes an observation $Z$ of length $n$, whose distribution is denoted by $\tilde{Q}^n$. For $j \in \llbracket 1, n \rrbracket$, we denote the distribution of each component $Z_j$ of $Z$ by $\tilde{Q}_j$, where

$$
\tilde{Q}_j(z) = \sum_{x_{(j)}[\mathcal{K}]} \prod_{k \in \mathcal{K}} \mu_k^{(n)}(x_{kj}z) W_{\mathcal{Z}[\mathcal{X}[\mathcal{K}]]}(z|x_{(j)}[\mathcal{K}])
$$

where (a) follows from the definition of $Q_T(z) \triangleq W_{\mathcal{Z}[\mathcal{X}[\mathcal{K}]]}(z|x_T)$ in (1). Alternatively, using Lemma 5 in the appendix, we write

$$
\tilde{Q}_j(z) = Q_{\theta}(z) + \sum_{T \subseteq \mathcal{K}; T \neq \emptyset} \left( \prod_{k \in T} \mu_k^{(n)} \right) G_T(z).
$$

From the definition of $\delta_n$, we have

$$
\delta_n = D(\tilde{Q}^n || Q_{\theta}^n)
$$

$$
= -H(Z) + E \tilde{Q}^n \left( \log \frac{1}{Q_{\theta}^n(Z)} \right)
$$

$$
= - \left( \sum_{j=1}^n H(Z_j|Z_{j-1}) \right) + E \tilde{Q}^n \left( \sum_{j=1}^n \log \frac{1}{Q_{\theta}(Z_j)} \right)
$$

$$
\geq \sum_{j=1}^n D(\tilde{Q}_j || Q_{\theta})
$$

$$
\lim_{n \to \infty} D(\tilde{Q}_j || Q_{\theta}) = 0,
$$

for all $j \in \llbracket 1, n \rrbracket$. Applying Pinsker’s inequality on (53), we obtain $D(\tilde{Q}_j || Q_{\theta}) = 0$, which implies that

$$
\forall z \in \mathcal{Z},
$$

$$
\lim_{n \to \infty} |\tilde{Q}_j(z) - Q_{\theta}(z)| = 0,
$$

$$
\lim_{n \to \infty} \tilde{Q}_j(z) = Q_{\theta}(z).
$$

Fixing $j = 1$ and by using (46) and (55), for $n \in \mathbb{N}^*$, we obtain (57) shown at the top of the next page. Since we assumed that the sequence $\{\mu_k^{(n)}\}_{k \in \mathcal{K}}$ is not an all-zero sequence, (57) implies that $Q_{\theta}$ is a convex combination of $\{Q_T\}_{T \subseteq \mathcal{K}; T \neq \emptyset}$. Note that the convex combination in (57) does not require the transmitters to coordinate, which is the case in our channel model, since the input from each user is independent of the inputs from other users. Since (57) contradicts the assumption made in Section III, our assumption about $\{\mu_k^{(n)}\}_{k \in \mathcal{K}}$ is incorrect, and we have

$$
\lim_{n \to \infty} \mu_k^{(n)} = 0,
$$

for all $k \in \mathcal{K}$, which implies that

$$
\lim_{n \to \infty} \mu_k^{(n)} = 0.
$$

Subsequently, from (43) and (59), we obtain

$$
\lim_{n \to \infty} \mu_k^{(n)} = 0,
$$

for all $(k, j) \in \mathcal{K} \times \llbracket 1, n \rrbracket$. Henceforth, we only consider the subsequence of codes with blocklength $n \in \mathbb{N}^*$. Next, for $j \in \llbracket 1, n \rrbracket$, define

$$
\psi_{j}^{(n)}(z) \triangleq |\tilde{Q}_j(z) - Q_{\theta}(z)|.
$$

Note that $\sum_j \psi_{j}^{(n)}(z) = 0$. Also note that from (54) and (61), we have $\lim_{n \to \infty} \psi_{j}^{(n)}(z) = 0$ for all $j \in \llbracket 1, n \rrbracket$ and $\forall z \in \mathcal{Z}$. We lower bound $D(\tilde{Q}_j || Q_{\theta})$ for $n$ large enough by (65) shown in the next page, where (a) follows from the inequality

$$
\log (1 + x) > x - \frac{x^2}{2} + \frac{x^3}{3} + \ldots
$$

for $x \geq 0$ and $4 \log (1 + x) > x - \frac{x^2}{2}$ for $x \in [-\frac{1}{2}, 0]$. For $j \in \llbracket 1, n \rrbracket$, define $\psi_{j}^{(n)}(z) \triangleq \frac{\psi_{j}^{(n)}(z)}{\psi_{j}^{(n)}(z) + \frac{1}{Q_{\theta}(z)} - 1}$ and $\xi_{j}^{(n)}(z) \triangleq \max_{z \in \mathcal{Z}} \psi_{j}^{(n)}(z)$. Since $\lim_{n \to \infty} \psi_{j}^{(n)}(z) = 0$, we have $\lim_{n \to \infty} \xi_{j}^{(n)}(z) = 0$ for all $j \in \llbracket 1, n \rrbracket$. From (47) and (61), for $j \in \llbracket 1, n \rrbracket$, we write

$$
\left| \psi_{j}^{(n)}(z) \right| = \left| \tilde{Q}_j(z) - Q_{\theta}(z) \right|
$$

$$
\leq \sum_{T \subseteq \mathcal{K}; T \neq \emptyset} \left( \prod_{k \in T} \mu_k^{(n)} \right) |G_T(z)|
$$

$$
\leq \mu_k^{(n)} \sum_{T \subseteq \mathcal{K}; T \neq \emptyset} |G_T(z)|
$$

where (a) follows from (43) and the fact that $\mu_k^{(n)} \in [0, 1]$ for all $k \in \mathcal{K}$ and $j \in \llbracket 1, n \rrbracket$. Note that the term inside the
\[
\lim_{n \to \infty} \left( \sum_{T \subseteq K} \left( \prod_{k \in T} \mu_{k1}^{(n)} \right) \left( \prod_{k \in T^c} \left( 1 - \mu_{k1}^{(n)} \right) \right) Q_T(z) \right) = \bar{Q}_0(z), \\
\sum_{T \subseteq K} \left( \prod_{k \in T} \mu_{k1}^{(n)} \right) \left( \prod_{k \in T^c} \left( 1 - \mu_{k1}^{(n)} \right) \right) Q_T(z) = \bar{Q}_0(z).
\]

\[
\mathbb{D}(\hat{Q}_j || \bar{Q}_0) = \sum_z \hat{Q}_j(z) \log \frac{\hat{Q}_j(z)}{\bar{Q}_0(z)}
\]

\[
= \sum_z \bar{Q}_0(z) \left( 1 + \frac{\Psi_j^{(n)}(z)}{\bar{Q}_0(z)} \right) \log \left( 1 + \frac{\Psi_j^{(n)}(z)}{\bar{Q}_0(z)} \right)
\]

\[
\geq \sum_z \left( \frac{(\Psi_j^{(n)}(z))^2}{2\bar{Q}_0(z)} - \frac{(\Psi_j^{(n)}(z))^3}{2\bar{Q}_0(z)^2} \right) + \sum_{z: \Psi_j^{(n)}(z) < 0} \frac{2(\Psi_j^{(n)}(z))^3}{3\bar{Q}_0(z)}
\]

\[
\geq \sum_z \frac{(\Psi_j^{(n)}(z))^2}{2\bar{Q}_0(z)} \left( 1 - \frac{\Psi_j^{(n)}(z)}{\bar{Q}_0(z)} - \frac{4|\Psi_j^{(n)}(z)|}{3\bar{Q}_0(z)} \right).
\]

**parentheses in (68) is positive and bounded. Consequently, for**

\[
\max_{j \in [1,n]} \left| \Psi_j^{(n)}(z) \right| \leq \mu_{k1}^{(n)} \left( \sum_{T \subseteq K: T \neq \emptyset} \left| G_T(z) \right| \right).
\]

**From the definition of \( \hat{\zeta}_j^{(n)}(z) \), we have**

\[
\hat{\zeta}_j^{(n)}(z) = \frac{\Psi_j^{(n)}(z)}{\bar{Q}_0(z)} + \frac{4|\Psi_j^{(n)}(z)|}{3\bar{Q}_0(z)} \]

\[
\leq \frac{\Psi_j^{(n)}(z)}{\bar{Q}_0(z)} + \frac{4|\Psi_j^{(n)}(z)|}{3\bar{Q}_0(z)}
\]

\[
= \frac{7|\Psi_j^{(n)}(z)|}{3\bar{Q}_0(z)}.
\]

**Consequently, we have**

\[
\hat{\zeta}^{(n)}(z) = \max_{j \in [1,n]} \hat{\zeta}_j^{(n)}(z)
\]

\[
\leq \frac{7}{3\bar{Q}_0(z)} \max_{j \in [1,n]} \left| \Psi_j^{(n)}(z) \right|
\]

\[
\leq \frac{7}{3\bar{Q}_0(z)} \mu_{k1}^{(n)} \left( \sum_{T \subseteq K: T \neq \emptyset} \left| G_T(z) \right| \right).
\]

**Note that, by definition, \( \hat{\zeta}_j^{(n)}(z) \) is non-negative irrespective of the sign of \( \Psi_j^{(n)}(z) \). Then, using (59) and (75), we conclude that**

\[
\lim_{n \to \infty} \hat{\zeta}^{(n)}(z) = 0.
\]

**Using (65), we lower bound (52) by**

\[
\delta_n \geq \sum_{j=1}^n \sum_z \left( \frac{(\Psi_j^{(n)}(z))^2}{2\bar{Q}_0(z)} \left( 1 - \hat{\zeta}_j^{(n)}(z) \right) \right)
\]

\[
\geq \sum_z \left( 1 - \hat{\zeta}_j^{(n)}(z) \right) \sum_{j=1}^n \left( \Psi_j^{(n)}(z) \right)^2.
\]

For \( k \in K \), we upper bound \( \log M_k \) using standard techniques,

\[
\log M_k \leq \mathbb{I}(W_k; Y) + \mathbb{H}_b(\epsilon_n) + \epsilon_n \log M_k
\]

\[
\leq \mathbb{I}(W_k; S_k) + \mathbb{H}_b(\epsilon_n) + \epsilon_n \log M_k
\]

\[
= \mathbb{I}(X_k; Y) + \mathbb{H}_b(\epsilon_n) + \epsilon_n \log M_k
\]

\[
= \mathbb{H}(X_k) - \mathbb{H}(X_k|Y) + \mathbb{H}_b(\epsilon_n) + \epsilon_n \log M_k
\]

\[
\leq \mathbb{H}(X_k|X[K\setminus\{k\}]) - \mathbb{H}(X_k|X[K]) + \mathbb{H}_b(\epsilon_n) + \epsilon_n \log M_k
\]

\[
\leq \sum_{j=1}^n \mathbb{H}(Y_j|X(j)\{k\}) - \sum_{j=1}^n \mathbb{H}(Y_j|X(j)|K)
\]

\[
= \sum_{j=1}^n \mathbb{H}(X_k; Y_j|X(j)|K) + \mathbb{H}_b(\epsilon_n)
\]

\[
+ \epsilon_n \log M_k,
\]

where \( (a) \) follows from Fano’s inequality, \( (b) \) follows from the fact that \( X_k \) and \( X[K\setminus\{k\}] \) are mutually independent and the fact that conditioning reduces entropy, and \( (c) \) follows from the fact that conditioning reduces entropy and the memoryless property of the channel. We expand the mutual information term in (68) as (88) shown at the top of the next page. Defining \( \mu_{\max}^{(n)} \triangleq \mu_{k1}^{(n)} \) and \( d_1 \triangleq 2^\mathbb{K} \max_{T \subseteq K: |T| > 1} \mathbb{D}(P_T || P_{\bar{0}}) \), we upper bound the first term in (88) by (90) shown in the next page, where \( (a) \) follows from splitting the sum into two based on the number of \( Y \)’s in \( x_T \), and \( (b) \) follows from the fact that \( 1 - \mu_{k1}^{(n)}(i, j) \leq 1 \) for all \( (i, j) \in K \times [1,n] \). Defining \( d_2 \triangleq 2^\mathbb{K} \max_{i \in K\setminus\{k\}} \mathbb{D}(W_{Y_j|X(i)|K\setminus\{k\}} = 0|X\{k\}|1) P_{\bar{0}} \), we lower bound the second term in (88) by (98) in the next page, where \( (a) \)
follows from the fact that \( \prod_{i \in \mathcal{K} \setminus \{k\}} \left( 1 - \mu_{ij}^{(n)} \right) = 1 + \sum_{T \subseteq \mathcal{K} \setminus \{k\}} (-1)^{|T|} \left( \prod_{i' \in T} \mu_{ij}^{(n)} \right) \). Note that we can write \( W_{Y_j|X_{ij}[\mathcal{K} \setminus \{k\}]}(y_{X_{ij}[\mathcal{K} \setminus \{k\}]}|y_{x_T[\mathcal{K} \setminus \{k\}]} = x_{T[\mathcal{K} \setminus \{k\}]}) = (1 - \mu_{ij}^{(n)}) P_i(y) + \mu_{ij}^{(n)} P_{i|k}(y) \). We define \( d_3 \triangleq \sum_y (P_{i|k}(y) - P_i(y)) \log \frac{P_i(y)}{P_{i|k}(y)} \). Note that \( d_3 \) is bounded since \( P_i \ll P_{i|k} \). Then, we lower bound the KL divergence term in (98) by (101). Defining \( d_4 \triangleq d_1 + d_2 + d_3 \) and combining (88), (90), (98), and (101), we obtain

\[
\mathbb{I}(X_{kj} ; Y_j|X_{ij}[\mathcal{K} \setminus \{k\}]) \leq \mu_{ij}^{(n)} \mathbb{D}(P_i \| P_{i|k}) + d_4 \mu_{ij}^{(n)} \sum_{i \in \mathcal{K} \setminus \{k\}} \mu_{ij}^{(n)} .
\]
\[
\mathbb{D} \left( W_{Y_j | X_{ij} | K \setminus k} \| P_0 \right) = \sum_y W_{Y_j | X_{ij} | K \setminus k} \left( y | x_j, K \setminus k \right) \log \frac{P_i(y)}{P_0(y)} + \mathbb{D} \left( W_{Y_j | X_{ij} | K \setminus k} = x_j, K \setminus k \right) \| P_i \right)
\]
\[
\geq \sum_y P_i(y) \left( 1 + \mu_{ij}^{(n)} \frac{P_i(k) - P_i(y)}{P_i(y)} \right) \log \frac{P_i(y)}{P_0(y)}
\]
\[
\geq \mathbb{D}(P_i \| P_0) - d_3 \mu_{\max}^{(n)}.
\]

\[
\log M_k \leq \frac{\sum_{j=1}^n \mu_{ij}^{(n)} \mathbb{D}(P_k \| P_0) + d_4 \mu_{\max}^{(n)} \sum_{i \in K} \sum_{j=1}^n \mu_{ij}^{(n)} + \mathbb{H}_b (\epsilon_n)}{(1 - \epsilon_n) \sqrt{n \sum_{z} \frac{(1 - \xi^{(n)}(z))}{2Q_b(z)} \sum_{j=1}^n \left( \psi_j^{(n)}(z) \right)^2}}
\]
\[
\leq \frac{\mathbb{D}(P_k \| P_0) \sum_{i \in K} \sum_{j=1}^n \mu_{ij}^{(n)} + d_4 \mu_{\max}^{(n)} + \sum_{i \in K} \sum_{j=1}^n \mu_{ij}^{(n)} + \mathbb{E}_b (\epsilon_n)}{(1 - \epsilon_n) \sqrt{n \sum_{z} \frac{(1 - \xi^{(n)}(z))}{2Q_b(z)} \sum_{j=1}^n \left( \psi_j^{(n)}(z) \right)^2}}
\]
\[
\mathbb{D}(P_k \| P_0) \mathbb{D}^{(a)}(P_k \| P_0) \sum_{i \in K} \sum_{j=1}^n \mu_{ij}^{(n)} + d_4 \mu_{\max}^{(n)} + \sum_{i \in K} \sum_{j=1}^n \mu_{ij}^{(n)} + \mathbb{E}_b (\epsilon_n)
\]
\[
(1 - \epsilon_n) \sqrt{n \sum_{z} \frac{(1 - \xi^{(n)}(z))}{2Q_b(z)} \sum_{j=1}^n \left( \psi_j^{(n)}(z) \right)^2}
\]

\[
\psi_j^{(n)}(z) = \sum_{i \in K} \mu_{ij}^{(n)} G_i(z) + \sum_{T \subseteq K, |T| \geq 2} \left( \prod_{k \in T} \mu_{kj}^{(n)} \right) G_T(z)
\]
\[
= \sum_{i \in K} \mu_{ij}^{(n)} G_i(z) + \sum_{T \subseteq K, i \in T, |T| \geq 2, \forall k \in T, k \neq i} \left( \prod_{k \in T \setminus \{i\} \subseteq K} \mu_{kj}^{(n)} \right) G_T(z)
\]
\[
= \sum_{i \in K} \mu_{ij}^{(n)} \left( G_i(z) + \sum_{T \subseteq K, i \in T, |T| \geq 2, \forall k \in T, k \neq i} \left( \prod_{k \in T \setminus \{i\} \subseteq K} \mu_{kj}^{(n)} \right) G_T(z) \right).
\]

Next, we normalize \( \log M_k \), where \( k \in K \), by \( \sqrt{n \epsilon_n} \). Using (77), (86), and (102), for \( n \) large enough, we obtain (106), where (a) follows from the fact that \( n \sum_{j=1}^n \left( \psi_j^{(n)}(z) \right)^2 \geq \left( \sum_{j=1}^n \psi_j^{(n)}(z) \right)^2 \) according to the Cauchy-Schwarz inequality. Note that since \( (1 - \xi^{(n)}(z)) \) is positive for \( n \) large enough, our application of Cauchy-Schwarz inequality in (106) is valid. From the definition of \( \psi_j^{(n)}(z) \) in (61), we have (109) shown at the top of this page. Define \( d_5 \triangleq 2^k \max_{z \in Z} \max_{T \subseteq K, |T| \geq 1} |G_T(z)| \). If \( \sum_{i \in K} \mu_{ij}^{(n)} G_i(z) \leq 0 \), we upper bound (109) by

\[
\psi_j^{(n)}(z) \leq \sum_{i \in K} \mu_{ij}^{(n)} \left( G_i(z) + d_5 \mu_{\max}^{(n)} \right),
\]
which is negative for \( n \) large enough. If \( \sum_{i \in K} \mu_{ij}^{(n)} G_i(z) \geq 0 \), we lower bound (109) by

\[
\psi_j^{(n)}(z) \geq \sum_{i \in K} \mu_{ij}^{(n)} \left( G_i(z) - d_5 \mu_{\max}^{(n)} \right),
\]
a convergent subsequence is bounded between 0 and 1 for any $a_{\mathbf{z}}$.

We only consider the subsequence of codes with blocklength $n$ large enough. Consequently, for $n$ large enough, combining (106), (110) and (111), we obtain (113) shown at the top of this page. Combining (86) and (102) with the fact that $\lim_{n \to \infty} \log M_k = \infty$, we conclude that

$$\lim_{n \to \infty} \sum_{i \in \mathcal{K}} \sum_{j=1}^{n} \| \mu_{ij}^{(n)} \|_{\mathcal{K}} = \infty.$$ 

Note that $\sum_{a \in \mathcal{K}} \rho_a \leq 1$. Since we have assumed in Section III that there exists no $\{\rho_k\}_{k \in \mathcal{K}}$ for which $\sum_{k \in \mathcal{K}} \rho_k \mathbb{Q}_k(z) = \mathbb{Q}_\emptyset(z)$ for all $z \in \mathcal{Z}$, the denominator in (113) is non-zero. Henceforth, we only consider the subsequence of codes with blocklength $n \in N^\updownarrow$. Defining $\rho_0 \triangleq \{\rho_k\}_{k \in \mathcal{K}}$, we obtain from (113),

$$\liminf_{n \to \infty} \frac{\log M_k}{\sqrt{n} \delta_n} \leq \sqrt{2} \rho_0 \mathbb{D}(P_k \| P_0)$$

where (a) follows from the definition of $\chi(\rho)$.

Using standard techniques, we lower bound $M_k L_k$, for $k \in \mathcal{K}$, by

$$\log M_k L_k \leq \mathbb{H}(W_k S_k)$$

where (a) follows from the fact that $X_k$ is a function of $W_k$ and $S_k$. Defining $d_6 \triangleq 2^k \max_{i \in \mathcal{K}} \mathbb{D}(Q_i \| Q_\emptyset)$, we then lower bound the first term in (119) by (125) shown at the bottom of this page, where (a) follows from the steps used to obtain (98) from (94). Note that, by definition, we have (126) shown in the next page. We upper bound the second term in (119) by (134), where (a) follows from the fact that conditioning reduces entropy and the memoryless property of the channel, and (b) follows from (126). Defining $d_7 \triangleq 2^k \max_{T \subseteq \mathcal{K} : |T| > |1} \mathbb{D}(Q_T \| Q_\emptyset)$, we upper bound the first term in (134) by

$$\sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \mu_{ij}^{(n)} \right) \mathbb{D}(Q_T \| Q_\emptyset)$$

where (a) follows from the definition of $\chi(\rho)$.
\[
\sum_{T \subseteq \mathcal{K} \setminus \{k\}} \left( \prod_{i \in \mathcal{K} \setminus \{k\}} \Pi x_{ij}(x_{T,i}) \right) W_{Z_j|X_{ij}[\mathcal{K} \setminus \{k\}]x_j}(z|x_T|\mathcal{K} \setminus \{k\})x = W_{Z_j|x_j}(z|x).
\] 
(126)

\[
\mathbb{I}(X|\mathcal{K} \setminus \{k\}; Z|X_k) \\
= \mathbb{H}(Z|X_k) - \mathbb{H}(Z|X[\mathcal{K}])
\]
(127)

\[
\leq \sum_{j=1}^{n} \left( \mathbb{H}(Z_j|X_{kj}) - \mathbb{H}(Z_j|X_{ij}[\mathcal{K}]) \right)
\]
(128)

\[
= \sum_{j=1}^{n} \mathbb{I}(X_{ij}[\mathcal{K} \setminus \{k\}]; Z_j|X_{kj})
\]
(129)

\[
= \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \mathbb{D}(Q_T\|W_{Z_j|x_{kj}}=x_{T,k})
\]
(130)

\[
= \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \mathbb{D}(Q_T\|Q_\emptyset) - \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \sum_{z} Q_T(z) \log \frac{W_{Z_j|x_{kj}}(z|x_{T,k})}{Q_\emptyset(z)}
\]
(131)

\[
= \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \mathbb{D}(Q_T\|Q_\emptyset) - \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \sum_{z} Q_T(z) \log \frac{W_{Z_j|x_{kj}}(z|x_{T,k})}{Q_\emptyset(z)}
\]
(132)

\[
\leq \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \mathbb{D}(Q_T\|Q_\emptyset) - \sum_{j=1}^{n} \mu^{(n)}_{kj} \mathbb{D}(W_{Z_j|x_{kj}}=x_i\|Q_\emptyset)
\]
(133)

\[
\leq \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in \mathcal{K}} \Pi x_{ij}(x_{T,i}) \right) \mathbb{D}(Q_T\|Q_\emptyset) - \sum_{j=1}^{n} \mu^{(n)}_{kj} \mathbb{D}(W_{Z_j|x_{kj}}=1\|Q_\emptyset).
\]
(134)

\[
\leq \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in T} \mu^{(n)}_{ij} \right) \mathbb{D}(Q_T\|Q_\emptyset)
\]
(135)

\[
= \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \mathbb{D}(Q_i\|Q_\emptyset) + \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \left( \prod_{i \in T} \mu^{(n)}_{ij} \right) \mathbb{D}(Q_T\|Q_\emptyset)
\]
(136)

\[
\leq \sum_{j=1}^{n} \sum_{T \subseteq \mathcal{K}} \mu^{(n)}_{ij} \mathbb{D}(Q_i\|Q_\emptyset) + d_t \mu^{(n)}_{ij} \max_{i \in \mathcal{K}} \sum_{j=1}^{n} \mu^{(n)}_{ij},
\]
(137)

where (a) follows from the fact that \( \left( \prod_{i \in T} (1 - \mu^{(n)}_{ij}) \right) \leq 1 \) for any \( T \subseteq \mathcal{K} \). Then, from Corollary 1, we write (139) as shown at the top of the next page. Defining

\[
d_8 \triangleq 2^K \max_{T \subseteq \mathcal{K} \setminus \{k\}: T \neq \emptyset} \left[ \sum_{z} \sum_{U \subseteq T} (-1)^{|T|-|U|} Q_U(x_{ij}|U) (z) \log \frac{Q_k(z)}{Q_\emptyset(z)} \right]
\]
(138)

and using (139), we bound the second KL divergence term in (134) by (143) shown in the next page. Defining \( d_9 \triangleq d_7 + d_8 \) and combining (134), (138), and (143), we obtain (145) shown in the next page. Defining \( d_{10} \triangleq d_6 + d_9 \) and combining (125) and (145), we bound (119) by

\[
\log M_k \geq \left( \sum_{j=1}^{n} \mu^{(n)}_{kj} \right) \mathbb{D}(Q_k\|Q_\emptyset) - d_{10} \mu^{(n)}_{kj} \sum_{i \in \mathcal{K}} \sum_{j=1}^{n} \mu^{(n)}_{ij} - \delta_n.
\]
(146)

Normalizing \( \log M_k L_k \), where \( k \in \mathcal{K} \), by \( \sqrt{n} \delta_n \), we obtain (148) as shown in the next page. Consider a sequence of codes for which (115) holds with equality for all \( k \in \mathcal{K} \). Proposition 1 confirms the existence of such schemes. As a result, for an arbitrary \( \zeta > 0 \), we have

\[
\liminf_{n \to \infty} \frac{\log M_k}{n \delta_n} \geq (1 - \zeta) \sqrt{\frac{2}{X(\rho)} \mathbb{D}(P_k\|P_\emptyset)}.
\]
(149)

Then, for that sequence of codes, using (86) and (102), we obtain (153) shown in the next page. However, since \( \limsup_{n \to \infty} a_n \geq \liminf_{n \to \infty} a_n \) for any sequence \( \{a_n\} \),
we write
\[ n \to \infty \limsup_{n \in \mathcal{N}^\dagger} \frac{\sum_{j=1}^n \mu_{kj}^{(n)}}{\sqrt{n \delta n}} \geq (1 - \zeta) \sqrt{\frac{2}{\chi(\rho)}} \rho_k. \] (154)

Combining (148) and (154), we obtain
\[ n \to \infty \limsup_{n \in \mathcal{N}^\dagger} \frac{\log M_k L_k}{\sqrt{n \delta n}} \geq (1 - \zeta) \sqrt{\frac{2}{\chi(\rho)}} \rho_k \] (155)

for an arbitrary \( \zeta > 0 \), where (a) follows from the fact that \( \lim_{n \to \infty} \sum_{j=1}^n \mu_{kj}^{(n)} = \infty \). Letting \( \zeta \downarrow 0 \) in (155), we obtain (41).

Note that for any sequence \( \{a_n\}_{n \in \mathbb{N}^*} \) and any infinite set \( \mathcal{N} \subseteq \mathbb{N}^* \), we have, by definition,
\[ n \to \infty \liminf_{n \in \mathcal{N}^\dagger} a_n \leq \liminf_{n \to \infty} a_n \leq \limsup_{n \in \mathcal{N}^\dagger} a_n \leq \limsup_{n \to \infty} a_n. \] (156)

From Proposition 2 and equation (156), we conclude that the covert capacity region is contained in the region defined
by (157) shown at the top of this page, and that, any achievable covert throughput tuple \( r \) is associated to an achievable key throughput of at least
\[
\frac{2}{\chi(\rho)} \rho_k \left( D(Q_k \| P) - D(P_k \| P_0) \right)_{+}^\top \text{ for each } k \in \mathcal{K}.
\]

VI. CONCLUSION

We conclude with a discussion of extensions of our results and related problems of interest.

Although we have limited our characterization of the covert capacity region to binary-input, K-user MACs, our results also extend to transmitters with non-binary input alphabets as in [5], [6]. In this case each user is characterized by a distinct input alphabet \( X_k \triangleq \{ x_i^{(k)} \}_{i=0} \) with one innocent symbol \( x_0^{(k)} \) and \( N_k \) information symbols. The steps required to analyze this setting are virtually identical to those presented in the binary case, but one must account for the possibility of optimizing a probability distribution over the information symbols. Upon denoting by \( P_{k,i} \) (resp. \( Q_{k,i} \)) the distribution induced at the output of the main (resp. warden) channel when user \( k \) transmits symbol \( x_i^{(k)} \) and all other users transmit their innocent symbol, the covert capacity region takes the form in (158) shown at the top of the next page with \( \chi(\rho, \beta) \) defined as in (159). We do not explicitly provide the detailed steps here in the interest of clarity and simplicity, since multiple information symbols are not as easily and concisely represented with our current notation for the one-shot output distributions at the receiver and the warden. The interested reader can refer to [39] for more details.

Our resolvability analysis is not directly applicable to AWGN channels since we use \( v_{\min} \triangleq \min_{n} Q_0(z) \) in the denominator of (258) and (291), which is zero for AWGN channels. However, our achievability results can be extended to AWGN channels by using resolvability exponents as in [40]--[42] to obtain a bound for the KL divergence \( D(Q^n \| Q^m_0) \) that does not rely on the discrete or continuous nature of the channel output alphabet. The converse argument can be developed by extending the approach of [5] to deal with multiple users, and one expects the covert capacity region to be
\[
\bigcup_{\{ r_k \}_{k \in \mathcal{K}} : \sum_k r_k = 1} \left\{ r_k \right\} r_k \leq \rho_k \right\}.
\]

A final problem of interest is the characterization of the covert capacity region of a K-user MAC in which the transmitters share a common key. Unlike the situation addressed here, the common key scenario captures the ability of users to coordinate their covert transmissions. One can approach the problem by following cooperative channel resolvability techniques studied in [43], [44].
\[
\bigcup_{(\varkappa)_{k \in \mathcal{K}} \in \{0, 1\}^\mathcal{K}, \sum_{k \in \mathcal{K}} \varkappa_k = 1}
\left\{r_k \in \mathcal{K} : \forall k \in \mathcal{K}, \ r_k \leq \sqrt{\frac{2}{\chi(\rho, \beta) N_k}} \sum_{i=1}^{N_k} \beta_{k,i} \mathbb{I}(P_{k,i} \| P) \right\}.
\]

(158)

\[
\chi(\rho, \beta) = \sum_z \frac{\left( \sum_{k \in \mathcal{K}} \rho_k \sum_{i \in [1, \beta]} \beta_{k,i} Q_k(z) - Q_0(z) \right)^2}{Q_0(z)}.
\]

(159)

\[
\prod_{k \in S'} (1 - \beta_k) = (1 - \beta_k) \prod_{k \in S} (1 - \beta_k)
\]

(164)

\[
= (1 - \beta_k) \left( 1 + \sum_{T \subseteq S: |T| \neq 0} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right) \right)
\]

(165)

\[
= 1 - \beta_k + \sum_{T \subseteq S: |T| \neq 0} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right) - \beta_k \left( \sum_{T \subseteq S: |T| \neq 0} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right) \right)
\]

(166)

\[
\equiv 1 + \sum_{T = (\mathcal{K})} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right) + \sum_{T \subseteq S: |T| = 1} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right) + \sum_{T \subseteq S: |T| > 1, \ k \in T} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right)
\]

(167)

\[
\equiv 1 + \sum_{T \subseteq S': |T| = 1} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right) + \sum_{T \subseteq S': |T| > 1, \ k \in T} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right)
\]

(168)

\[
= 1 + \sum_{T \subseteq S': |T| > 1} (-1)^{|T|} \left( \prod_{k \in T} \beta_k \right)
\]

(169)

\[
Q_{\alpha_n}(z) = \sum_{X_k} \left( \prod_{k \in \mathcal{K}} \Pi_{X_k}(x_k) \right) W_{Z|X}[\mathcal{K}|z|\mathcal{K}] \]

(171)

\[
= \sum_{T \subseteq \mathcal{K}} \left( \prod_{k \in T} \rho_k \alpha_n \right) \left( \prod_{k \in T^c} (1 - \rho_k \alpha_n) \right) Q_T(z)
\]

(172)

\[
= \left( \prod_{k \in \mathcal{K}} (1 - \rho_k \alpha_n) \right) Q_0(z) + \sum_{T \subseteq \mathcal{K}: T \neq \emptyset} \left( \prod_{k \in T} \rho_k \alpha_n \right) \left( \prod_{k \in T^c} (1 - \rho_k \alpha_n) \right) Q_T(z)
\]

(173)

\[
\equiv \left( \prod_{k \in \mathcal{K}} (1 - \rho_k \alpha_n) \right) Q_0(z) + \sum_{T \subseteq \mathcal{K}: T \neq \emptyset} \left( \prod_{k \in T} \rho_k \alpha_n \right) \left( \sum_{U \subseteq T^c: U \neq \emptyset} (-1)^{|U|} \left( \prod_{k \in U} \rho_k \alpha_n \right) \right) Q_T(z)
\]

(174)

\[
\equiv \left( \prod_{k \in \mathcal{K}} (1 - \rho_k \alpha_n) \right) Q_0(z) + \sum_{T \subseteq \mathcal{K}: T \neq \emptyset} \left( \prod_{k \in T} \rho_k \alpha_n \right) \left( \sum_{U \subseteq T^c} (-1)^{|U|} \left( \prod_{k \in U} \rho_k \alpha_n \right) \right) Q_T(z).
\]

(175)
\( Q_{a_n}(z) = \left( \prod_{k \in K} (1 - \rho_k a_n) \right) Q_\theta(z) + \sum_{T \subseteq K, T \neq \emptyset} \left( \prod_{k \in T} \rho_k a_n \right) \left( \sum_{S \subseteq K \setminus T} (-1)^{|S| - |T|} \left( \prod_{k \in (S \setminus T)} \rho_k a_n \right) \right) Q_T(z) \)  

(176)

\( = Q_\theta(z) + \sum_{S \subseteq K : S \neq \emptyset} \left( -1 \right)^{|S|} \left( \prod_{k \in S} \rho_k a_n \right) Q_\theta(z) + \sum_{T \subseteq K, T \neq \emptyset} \left( \sum_{S \subseteq K : S \neq \emptyset} (-1)^{|S| - |T|} \left( \prod_{k \in S} \rho_k a_n \right) \right) Q_T(z) \)

(177)

\( = Q_\theta(z) + \sum_{S \subseteq K : S \neq \emptyset} \left( \prod_{k \in S} \rho_k a_n \right) \left( \sum_{T \subseteq S : T \neq \emptyset} (-1)^{|S| - |T|} Q_T(z) \right) \)  

(178)

\( = Q_\theta(z) + \sum_{S \subseteq K : S \neq \emptyset} \left( \prod_{k \in S} \rho_k a_n \right) \left( \sum_{T \subseteq S} (-1)^{|S| - |T|} Q_T(z) \right) \)

(179)

\[ W_{Z|X_k}(z|1) = Q_k(z) + \sum_{S \subseteq K \setminus \{k\} : S \neq \emptyset} \left( \prod_{k \in S} \rho_k a_n \right) \left( \sum_{T \subseteq S} (-1)^{|S| - |T|} Q_T(z) \right) \]  

(181)

**APPENDIX B**

**PROOF OF LEMMA 1**

From the definition of \( \mathbb{D}(Q_{a_n} \parallel Q_\theta) \), we have

\[ \mathbb{D}(Q_{a_n} \parallel Q_\theta) = \sum_z Q_{a_n}(z) \log \frac{Q_{a_n}(z)}{Q_\theta(z)} \]

(182)

\[ = \sum_z Q_\theta(z) \left( 1 + \frac{a_n \zeta_n(z)}{Q_\theta(z)} \right) \log \left( 1 + \frac{a_n \zeta_n(z)}{Q_\theta(z)} \right) \]  

(183)

Since \( \log(1 + x) < x - \frac{x^2}{2} + \frac{x^3}{3} \), for \( x > -1 \), we upper bound (183) by

\[ \mathbb{D}(Q_{a_n} \parallel Q_\theta) \leq \sum_z Q_\theta(z) \left( 1 + \frac{a_n \zeta_n(z)}{Q_\theta(z)} \right) \times \left( \frac{a_n \zeta_n(z)}{Q_\theta(z)} - \frac{a_n^2 \zeta_n^2(z)}{2Q_\theta(z)} + \frac{a_n^3 \zeta_n^3(z)}{3Q_\theta(z)} \right) \]  

(184)

\[ = \sum_z \frac{a_n^2}{2} \left( \frac{\zeta_n^2(z)}{Q_\theta(z)} - \frac{a_n \zeta_n(z)}{3Q_\theta(z)} + \frac{2a_n^2 \zeta_n^4(z)}{3Q_\theta(z)} \right) \]  

(185)

where, \( (a) \) follows from the fact that \( \sum_z \zeta_n(z) = 0 \) from the definition of \( \zeta_n \). Since \( \lim_{n \to \infty} a_n = 0 \), \( a_n \) is small enough for a sufficiently large \( n \) and \( \frac{a_n \zeta_n(z)}{Q_\theta(z)} \in \left[ -\frac{1}{2}, 0 \right] \) for any \( z \in \mathcal{Z} \) if \( \zeta_n(z) < 0 \). Then, we lower bound (183) by (187), where, \( (a) \) follows from the inequalities \( \log(1 + x) > x - \frac{x^2}{2} \) for \( x \geq 0 \) and \( \log(1 + x) < x - \frac{x^2}{2} + \frac{x^3}{3} \) for \( x \in \left[ -\frac{1}{2}, 0 \right] \), and \( (b) \) follows from the fact that \( \sum_z \zeta_n(z) = 0 \). For \( n \) large enough, we loosen the bounds in (185) and (187) to obtain

\[ \frac{a_n^2}{2} \left( 1 + \sqrt{a_n} \right) \chi_n(\rho) \geq \mathbb{D}(Q_{a_n} \parallel Q_\theta) \geq \frac{a_n^2}{2} \left( 1 - \sqrt{a_n} \right) \chi_n(\rho) \]  

(188)

Finally, for a non-empty set \( T \subseteq K \), define \( \lambda_{n,T}(z) = \frac{\mathbb{D}(W_{Z|X_k}(z|1) - Q_\theta(z))}{Q_\theta(z)} \), Note that \( \sum_z \lambda_{n,T}(z) = 0 \). Then, for any non-empty set \( T \subseteq K \), we have (203) shown in the next page, where \( (a) \) follows from splitting the first term in (196) into three based on the number of users sending symbol 1, and \( (b) \) follows from the fact that the second term in (198) can be reduced to \( O(a_n^2) \) by expanding \( \log \left( 1 + a_n \frac{\lambda_{n,T}(z)}{Q_\theta(z)} \right) \) using Taylor series.
Lemma 6. Let \( \{ U_i \}_{i=1}^n \) be independent zero-mean random variables such that \( |U_i| \leq c \) for a finite \( c > 0 \) almost surely for all \( i \in [1, n] \). Then, for any \( t > 0 \),

\[
P\left( \sum_{i=1}^n U_i > t \right) \leq \exp\left( -\frac{t^2}{2 \sum_{i=1}^n \mathbb{E}(U_i^2) + \frac{1}{2}ct} \right).
\]  

(204)

APPENDIX C  
BERNSTEIN’S INEQUALITY

\[
\mathbb{I}(X[T]; Z) = \sum_{x \in \mathcal{X}[T]} \sum_{z \in \mathcal{Z}} \left( \prod_{k \in \mathcal{T}} \Pi_{x_k} (x_k) \right) W_{Z|X[T]} (z|x[T]) \log \left( \frac{W_{Z|X[T]} (z|x[T])}{Q_{a_n} (z)} \right)
\]

(195)

\[
= \sum_{x \in \mathcal{X}[T]} \sum_{z \in \mathcal{Z}} \left( \prod_{k \in \mathcal{T}} \Pi_{x_k} (x_k) \right) W_{Z|X[T]} (z|x[T]) \log \left( \frac{W_{Z|X[T]} (z|x[T])}{Q_\theta (z)} \right) - \mathbb{D}(Q_{a_n} \| Q_\theta)
\]

(196)

\[
\geq \sum_{k \in \mathcal{T}} \sum_{z \in \mathcal{Z}} \left( \prod_{k \in \mathcal{T}} (1 - \rho_k a_n) \right) W_{Z|X[T]} (z|x_k[T]) \log \left( \frac{W_{Z|X[T]} (z|x_k[T])}{Q_\theta (z)} \right)
\]

(197)

\[
\geq \sum_{k \in \mathcal{T}} \rho_k a_n \left( \sum_{x \in \mathcal{X}} \prod_{k \in \mathcal{T}} (1 - \rho_k a_n) \right) W_{Z|X[T]} (z|x_k[T]) \log \left( \frac{W_{Z|X[T]} (z|x_k[T])}{Q_\theta (z)} \right)
\]

(198)

\[
\geq \sum_{k \in \mathcal{T}} \rho_k a_n \sum_{z} W_{Z|X[T]} (z|x_k[T]) \log \left( \frac{W_{Z|X[T]} (z|x_k[T])}{Q_\theta (z)} \right) - \mathbb{D}(Q_{a_n} \| Q_\theta) + \mathcal{O} \left( a_n^2 \right)
\]

(199)

\[
\mathbb{D}(Q_{a_n} \| Q_\theta) + \mathcal{O} \left( a_n^2 \right)
\]

(200)

\[
\mathbb{D}(Q_k \| Q_\theta) + \sum_{z} Q_k(z) \log \left( \frac{Q_k(z)}{Q_\theta(z)} \right) - \mathbb{D}(Q_{a_n} \| Q_\theta) + \mathcal{O} \left( a_n^2 \right)
\]

(201)

\[
\mathbb{D}(Q_{a_n} \| Q_\theta) + \sum_{z} Q_k(z) \log \left( \frac{Q_k(z)}{Q_\theta(z)} \right) - \mathbb{D}(Q_{a_n} \| Q_\theta) + \mathcal{O} \left( a_n^2 \right)
\]

(202)

\[
\sum_{k \in \mathcal{T}} \rho_k a_n \mathbb{D}(Q_k \| Q_\theta) - \mathbb{D}(Q_{a_n} \| Q_\theta) + \mathcal{O} \left( a_n^2 \right)
\]

(203)

APPENDIX D  
PROOF OF LEMMA 2

The \( K \) users encode messages \( W[K] = m[K] \) using keys \( S[K] = \ell[K] \) into codewords \( x_K (m[K], \ell[K]) \) and transmit them over a discrete memoryless MAC. The following two events lead to a decoding error:

- The transmitted codewords do not satisfy \( (x_K (m[K], \ell[K]), y) \in \mathcal{A}_T^n \).
\[
E(P^m) = P(\hat{W}[K] \neq W[K])
\]

\[
= E \left( \frac{1}{(\prod_{k \in K} M_k)} \sum_{m[K]} \sum_{y} W^m_{Y|X[K]}(y|X_K(m[K], \ell[K])) I \left\{ \mathcal{E}_{m[K]} \cup \bigcup_{\tilde{m}[K] \neq m[K]} \mathcal{E}_{\tilde{m}[K]} \right\} \right)
\]

(206)

\[
\leq a \cdot E \left( \frac{1}{(\prod_{k \in K} M_k)} \sum_{m[K]} \sum_{y} W^m_{Y|X[K]}(y|X_K(m[K], \ell[K])) I \left\{ \mathcal{E}_{m[K]} \right\} \right) + E \left( \frac{1}{(\prod_{k \in K} M_k)} \sum_{m[K]} \sum_{y} W^m_{Y|X[K]}(y|X_K(m[K], \ell[K])) I \left\{ \mathcal{E}_{m[K]} \right\} \right)
\]

(207)

(208)

\[
E \left( \frac{1}{(\prod_{k \in K} M_k)} \sum_{m[K]} \sum_{y} W^m_{Y|X[K]}(y|X_K(m[K], \ell[K])) I \left\{ \mathcal{E}_{m[K]} \right\} \right)
\]

\[
= \sum_{x[K]} \sum_{y} W^m_{Y|X[K]}(y|X_K) \left( \prod_{k \in K} \Pi_{X_k}(x_k) \right) I \left\{ (x[K], y) \in \mathcal{A}^n_{\gamma} \right\}
\]

(209)

\[
= \sum_{T \subseteq K: T \neq \emptyset} \sum_{\gamma \in T} \left( A^n_{\gamma} \right)
\]

(210)

\[
\leq a \cdot \sum_{T \subseteq K: T \neq \emptyset} \left( A^n_{\gamma} \right)
\]

(211)

\[
= \sum_{T \subseteq K: T \neq \emptyset} \sum_{i=1}^{\gamma_{T \subseteq K}} \log \left( \frac{W_{Y|X[K]}(Y|X[K])}{W_{Y|X[T]}(Y|X[T])} \right) < \gamma_T
\]

(212)

(213)

- A different message vector \( \tilde{m}[K] \neq m[K] \) exists such that \( (x[K], \tilde{m}[K], \ell[K]), y) \in \mathcal{A}^n_{\gamma} \).

Define the event

\[
\mathcal{E}_{m[K]} \triangleq \left\{ (X_K(m[K], \ell[K]), Y) \in \mathcal{A}^n_{\gamma} \right\}.
\]

(205)

The probability of decoding error at the legitimate receiver averaged over all random codebooks is given by (208) at the top of this page, where \( (a) \) follows from the application of the union bound. We bound the first term in (208) by (213) shown at the top of this page, where \( (a) \) follows from the fact that \( \mathcal{A}^n_{\gamma} = \bigcup_{T \subseteq K: T \neq \emptyset} \mathcal{A}^n_{\gamma_T} \) and the application of the union bound. We define a zero-mean\(^5\) random variable \( U_T \triangleq I(X[T]; Y[X[T]]) - \log \left( W_{Y[X[T]]}(Y[X[T]]) \right) \). Note that \( |U_T| \) is bounded almost surely, and

\[
E(U_T^2) = E \left( \log^2 \left( \frac{W_{Y[X[T]]}(Y[X[K]])}{W_{Y[X[T]]}(Y[X[T])}} \right) - I(X[T]; Y[X[T]])^2 \right).
\]

(214)

Analyzing the expectation term on the right hand side of (214), we obtain (218) in the next page, where \( (a) \) follows from splitting the first sum on the right hand side of (215) into two based on whether \( x[K] \) equals \( x_0[K] \) or not, \( (b) \) follows from the fact that the first term in (216) is on the order of \( a_n \) since at least one of the symbols in \( x[K] \) is a 1, and \( (c) \) follows from the expansion of the product term and the fact that \( \log^2 \left( \frac{W_{Y[X[T]]}(Y[X[K]])}{W_{Y[X[T]]}(Y[X[T])}} \right) = \frac{\log^2 \left( W_{Y[X[T]]}(Y[X[K]]) \right)}{P_0(\gamma)} \). Expanding the numerator in the log^2 term in (218), we obtain (221), where \( (a) \) follows from the fact that the first term in (220) is on the order of \( a_n \) since at least one of the symbols in \( x[T] \) is a 1 and from the expansion of the product term. Combining (218) and (221), we obtain (223), where \( (a) \) follows from using the Taylor series of the log term. Let us now analyze the mutual information term on the right hand side of (214) as shown in (226), where \( (a) \) follows from Lemma 1. Using the definition of \( \gamma_T \), for an arbitrary \( \mu \in (0,1) \), we upper bound (213) using Bernstein’s inequality as shown in (231) for appropriate constants \( c, c_1, c_2 > 0 \), where \( (a) \) follows from using Bernstein’s inequality, and \( (b) \) follows from the fact that, for a finite \( K \), there exist \( 2^K - 1 \) non-empty subsets.
for any non-empty set of \( m \) left hand side of (232) in terms of the positions in which (234) is a subset of \( c \in E(\delta) \leq \log_2 T \). Combining (231) and (237), we upper bound the second term in (208) by (237), where (a) follows from rewriting the left hand side of (232) in terms of the positions in which the two vectors \( m[K] \) and \( \bar{m}[K] \) do not match, (b) follows from setting \( m[K] = 1[K] \) without loss of generality, and (c) follows from the fact that \( A^\alpha_n \) in the indicator function of (234) is a subset of \( A_{\gamma_T}^\alpha \) in the indicator function of (235) by definition of \( A_{\gamma_T}^\alpha \). Combining (231) and (237), we upper bound (208) by

\[
\mathbb{E} \left( \log^2 \frac{W_{Y[X[K]}(Y[X[K])}{W_{Y[X[T^c]}(Y[X[T^c])}} \right) = \sum_y \sum_{x[K]} \left( \prod_{k \in K} \pi_{X_k}(x_k) \right) W_{Y[X[K]}(y|x[K]) \log^2 \frac{W_{Y[X[K]}(y|x[K])}{W_{Y[X[T^c]}(y|x[T^c])}}
\]

Using the definition of \( \gamma_T \), (226) and (238), we conclude that for an arbitrary \( \delta \in (0, 1) \) and \( n \) large enough, if

\[
\sum_{k \in T} \log M_k = (1 - \delta)(1 - \mu)n(a_n \sum_{k \in T} \rho_k \| P_k \| P_\theta) + O(a_n^2)
\]

for every non-empty set \( T \subseteq K \), then there exists a constant \( \xi > 0 \) such that

\[
\mathbb{E}(P_n^\alpha) \leq \exp(-\xi n a_n).
\]

If \( T \) is a singleton set \( \{k\} \), where \( k \in K \), it follows from (239) that

\[
\log M_k = (1 - \delta)(1 - \mu)\rho_k n(a_n \sum_{k \in T} \rho_k \| P_k \| P_\theta).
\]

Observing (239) and (241), we conclude that (239) is automatically satisfied for every non-empty set \( T \subseteq K \), if \( \log M_k \) satisfies (241) for every \( k \in K \).
\[
\begin{align*}
    &\mathbb{E}\left(\frac{1}{(\prod_{k \in \mathcal{K}} M_k)^m} \sum_{y} W^n_{Y_{\mathcal{K}}|X_{\mathcal{K}}}(y|X_{\mathcal{K}}(m[\mathcal{K}]), \ell([\mathcal{K}])) \mathbbm{1}\{E_{\mathcal{m}[\mathcal{K}]}\} \right) \\
    \leq &\sum_{T \subseteq \mathcal{K}: T \neq \emptyset} \mathbb{P}\left(\sum_{i=1}^{n} U_T > \mu \bar{\eta}(X[T]; Y[X[T^c]])\right) \\
    = &\sum_{T \subseteq \mathcal{K}: T \neq \emptyset} \mathbb{P}\left(\sum_{i=1}^{n} U_T > \mu \bar{\eta}(X[T]; Y[X[T^c]])\right) \\
    \overset{(a)}{\leq} &\sum_{T \subseteq \mathcal{K}: T \neq \emptyset} \exp\left(-\frac{1}{\ell} \left(\mu \bar{\eta}(X[T]; Y[X[T^c]])\right)^2\right) \\
    \overset{(b)}{\leq} &\exp(-c_1 n a_n) \\
    \overset{(c)}{\leq} &\exp(-c_2 n a_n).
\end{align*}
\]
\[
\mathbb{E}(\mathbb{D}(\hat{Q}^n \parallel Q_{a_n})) = \mathbb{E}\left(\sum_z \hat{Q}^n(z) \log \frac{\hat{Q}^n(z)}{Q_{a_n}(z)}\right)
\]
\[
= \mathbb{E}\left(\sum_z \sum_{[m] \in \mathcal{E}(\mathcal{X}^T)} \frac{W^n_{Z|X}[z|x_T(m[T], \ell([\mathcal{K}])]}{(\prod_{k \in \mathcal{K}} M_k L_k)} \log \left(\frac{\sum_{[m] \in \mathcal{E}(\mathcal{X}^T)} W^n_{Z|X}[z|x_T(m[T], \ell([\mathcal{K}])]}{(\prod_{k \in \mathcal{K}} M_k L_k)} Q_{a_n}(z)\right)\right)
\]
\[
= \mathbb{E}\left(\sum_z \sum_{m[K] \in \mathcal{E}(\mathcal{X}^T)} W^n_{Z|X}[z|x_T(m[K], \ell([\mathcal{K}])]} \left(\prod_{k \in \mathcal{K}} M_k L_k\right) Q_{a_n}(z)\right)
\]
\[
\times \log \mathbb{E}_{\sim([m], \ell([\mathcal{K}])}(\sum_{[m] \in \mathcal{E}(\mathcal{X}^T)} W^n_{Z|X}[z|x_T(m[K], \ell([\mathcal{K}])]} \left(\prod_{k \in \mathcal{K}} M_k L_k\right) Q_{a_n}(z)).
\] 

(244)

(245)

(246)

\textbf{APPENDIX E}

\textbf{PROOF OF LEMMA 3}

Define the set \( B^T_{\eta} \triangleq \bigcap_{T' \neq \emptyset} B^T_{\eta_{T'}^T} \) with

\[
\text{\( B^T_{\eta_{T'}} \triangleq (x[T], z) \in \mathcal{A}^T \times \mathcal{Z}^T : \log \frac{W^n_{Z|X}[z|x_T]}{Q_{a_n}(z)} \leq \eta_{T'} \),}
\]

which follows from the fact that we can upper bound both \( \frac{W^n_{Z|X}[z|x_T(m[T], \ell(T))]}{(\prod_{k \in \mathcal{K}} M_k L_k)} \) and \( Q_{a_n}(z) \) by 1 and (b) follows from the fact that there only exist \( 2^K - 1 \) non-empty subsets of \( \mathcal{K} \). Combining (254) and (258), we upper bound (246) by (260) shown on page 23. From the definition of \( B^T_{\eta} \), we obtain

\[
\mathbb{P}(B^T_{\eta_{T'}}) \leq \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \mathbb{P}(B^T_{\eta_{T'}}),
\]

(261)

(262)

where (a) follows from Lemma 1. Let us analyze the expectation term in (263) as shown in (267) at the top of page 23, where (a) follows from the fact that \( B^T_{\eta_{T'}} \) is bounded almost surely, and

\[
\mathbb{E}(V_{T}^2) \leq \mathbb{E}\left(\frac{\log^2 \frac{W^n_{Z|X}[z|x_T]}{Q_{a_n}(z)}}{Q_{a_n}(z)} - (\mathbb{I}(X[T]; Z))^2 + c\right) + o(a_n^2),
\]

(263)

Defining \( \nu_{\min} \triangleq \min_{z} Q_{\hat{y}}(z) \), we upper bound the second term in (251) by (258) shown on page 23, where (a) follows from the fact that we can upper bound both \( \frac{\log^2}{Q_{a_n}(z)} \) and \( Q_{a_n}(z) \) by 1 and (b) follows from the fact that there only exist \( 2^K - 1 \) non-empty subsets of \( \mathcal{K} \). Combining (254) and (258), we upper bound (246) by (260) shown on page 23. From the definition of \( B^T_{\eta} \), we obtain

\[
Q_{a_n}(z) = \sum_{x[T]} W_{Z|X}[z|x_T] \left(\prod_{k \in \mathcal{T}} \Pi x_k(x_k)\right)
\]

(268)

(269)

(270)
Combining (267) and (270), we obtain

\[
\log E_{\sim(m[K], \ell[K])} \left( \frac{\sum_{T \subseteq K} \sum_{\bar{m}[K]} W_{Z|X[K]}^{zn}(z|x_K(\bar{m}[K], \bar{\ell}[K]))}{(\prod_{k \in K} M_k L_k) \cdot Q_{\alpha_n}(z)} \right) 
\]

\[
= \log \left( \frac{\sum_{T \subseteq K} \sum_{\bar{m}[T]} \sum_{\bar{\ell}[T]} W_{Z|X[T]}^{zn}(z|x_T(m[T], \ell[T]))}{(\prod_{k \in T} M_k L_k) \cdot Q_{\alpha_n}(z)} \right) 
\]

\[
= \log \left( \frac{\sum_{T \subseteq K} \sum_{\bar{m}[T]} \sum_{\bar{\ell}[T]} W_{Z|X[T]}(z|x_T(m[T], \ell[T]))}{(\prod_{k \in T} M_k L_k) \cdot Q_{\alpha_n}(z)} \right) + O(\alpha_n) 
\]

\[
\leq \log \left( \frac{\sum_{T \subseteq K} \sum_{\bar{m}[T]} \sum_{\bar{\ell}[T]} W_{Z|X[T]}(z|x_T(m[T], \ell[T]))}{(\prod_{k \in T} M_k L_k) \cdot Q_{\alpha_n}(z)} \right) + 1 \cdot 1 \left\{ \{x_K(m[K], \ell[K]), z\} \in B_{\eta} \right\} 
\]

\[
+ \log \left( \frac{\sum_{T \subseteq K} \sum_{\bar{m}[T]} \sum_{\bar{\ell}[T]} W_{Z|X[T]}(z|x_T(m[T], \ell[T]))}{(\prod_{k \in T} M_k L_k) \cdot Q_{\alpha_n}(z)} \right) + 1 \cdot 1 \left\{ \{x_K(m[K], \ell[K]), z\} \notin B_{\eta} \right\} 
\]

\[
\sum_{x_K(\bar{m}[K], \bar{\ell}[K])} W_{Z|X[K]}(z|x_K(\bar{m}[K], \bar{\ell}[K])) \left( \prod_{k \in K} \Pi_{x_k}^{\alpha_n}(x_k(\bar{m}[K], \bar{\ell}[K])) \right) = Q_{\alpha_n}^{zn}(z). 
\]

arbitrary \( \mu > 0 \), we upper bound (261) by

\[
\mathbb{E} \left( \log^2 \frac{W_{Z|X[T]}(Z|X[T])}{Q_{\alpha_n}(Z)} \right) 
\]

\[
= \sum_z W_{Z|X[T]}(z|x_0[T]) \log^2 \left( 1 + \frac{O(\alpha_n)}{W_{Z|X[T]}(z|x_0[T])} \right) 
\]

\[
\overset{(a)}{=} O(\alpha_n), 
\]

where (a) follows from the application of the Taylor series of the log term. Using the definition of \( \eta_T \) in (243), for an
\[
\log \left( \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \frac{W_{Z|X[T]}(z|x_T(m[T], \ell[T]))}{\left( \prod_{k \in T} M_k L_k \right) Q_{a_n}(z)} + 1 \right) \mathbb{I} \left\{ (\mathbf{x}_\mathcal{K}(m[\mathcal{K}], \ell[\mathcal{K}]), z) \notin \mathcal{B}_n^c \right\} \\
\leq \left( \log \left( \frac{1}{Q_{a_n}(z)} \right) + \log \left( \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \frac{W_{Z|X[T]}(z|x_T(m[T], \ell[T]))}{\left( \prod_{k \in T} M_k L_k \right) + Q_{a_n}(z)} \right) \right) \mathbb{I} \left\{ (\mathbf{x}_\mathcal{K}(m[\mathcal{K}], \ell[\mathcal{K}]), z) \notin \mathcal{B}_n^c \right\} \\
(255)
\]

\[
E(\mathbb{D}(\hat{Q}_n \parallel Q_{a_n})) \leq \sum_z Q_{a_n}(z) \left( \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \frac{e^{\eta T}}{\left( \prod_{k \in T} M_k L_k \right)} \right) + n \log \left( \frac{2^K}{\prod_{k \in \mathcal{K} : T \neq \emptyset} (1 - \rho_k a_n) \nu_{\text{min}}} \right) \mathbb{P}(B_n^c) \\
= \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \frac{e^{\eta T}}{\left( \prod_{k \in T} M_k L_k \right)} + n \log \left( \frac{2^K}{\prod_{k \in \mathcal{K}} (1 - \rho_k a_n) \nu_{\text{min}}} \right) \mathbb{P}(B_n^c). \\
(259)
\]

\[
E\left( \log^2 \frac{W_{Z|X[T]}(z|x_T[T])}{Q_{a_n}(z)} \right) = \sum_z \sum_{x[T]} \left( \prod_{k \in T} \Pi_{X_k}(x_k) \right) W_{Z|X[T]}(z|x_T[T]) \log^2 \frac{W_{Z|X[T]}(z|x_T[T])}{Q_{a_n}(z)} \\
\overset{(a)}{=} \sum_z \left( \prod_{k \in T} (1 - \rho_k a_n) \right) W_{Z|X[T]}(z|x_0[T]) \log^2 \frac{W_{Z|X[T]}(z|x_0[T])}{Q_{a_n}(z)} \\
+ \sum_z \sum_{x[T] \neq x[T]} \left( \prod_{k \in T} \Pi_{X_k}(x_k) \right) W_{Z|X[T]}(z|x_T[T]) \log^2 \frac{W_{Z|X[T]}(z|x_T[T])}{Q_{a_n}(z)} \\
\overset{(b)}{=} \sum_z \left( \prod_{k \in T} (1 - \rho_k a_n) \right) W_{Z|X[T]}(z|x_0[T]) \log^2 \frac{W_{Z|X[T]}(z|x_0[T])}{Q_{a_n}(z)} + \mathcal{O}(a_n) \\
\overset{(c)}{=} \sum_z W_{Z|X[T]}(z|x_0[T]) \log^2 \frac{Q_{a_n}(z)}{W_{Z|X[T]}(z|x_0[T])} + \mathcal{O}(a_n). \\
(264)
\]

\[
\overset{(a)}{\leq} \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \exp \left( -\frac{1}{n} \frac{(\mu n \Pi(X[T]; Z))^2}{\mathcal{O}(a_n)} + \frac{c_3 \mu n \Pi(X[T]; Z)}{2} \right) \\
\overset{(b)}{\leq} \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \exp(-c_1 na_n) \\
\overset{(c)}{\leq} \exp(-c_2 na_n), \\
(267)
\]

for appropriate constants \(c, c_1, c_2 > 0\), where \((a)\) follows from using Bernstein's inequality, and \((b)\) follows from the fact that \(\Pi(X[T]; Z) = \sum_{k \in T} \rho_k a_n \mathbb{D}(Q_k \parallel Q_0) + \mathcal{O}(a_n^2)\), for any non-empty set \(T \subseteq \mathcal{K}\), from Lemma 1. Combining \(260\) and \(278\), for an appropriate constant \(c_3 > 0\), we obtain

\[
E(\mathbb{D}(\hat{Q}_n^a \parallel Q_{a_n})) \leq \sum_{T \subseteq \mathcal{K} : T \neq \emptyset} \frac{e^{\eta T}}{\left( \prod_{k \in T} M_k L_k \right)} + \exp(-c_3 na_n). \\
(276)
\]

Using the definition of \(\eta_T\), we conclude from \(279\) that for an arbitrary \(\delta \in (0, 1)\) and a large \(n\), if

\[
\sum_{k \in T} \log(M_k L_k) = (1 + \delta) (1 + \mu) na_n \sum_{k \in T} \rho_k \mathbb{D}(Q_k \parallel Q_0), \\
(280)
\]
\[
\sum_{z} \left( \hat{Q}^n(z) - Q_{\bar{a}_n}^n(z) \right) \log \left( \frac{Q_{\bar{a}_n}^n(z)}{Q_{\bar{b}}^n(z)} \right) \\
\leq \sum_{z} \left| \hat{Q}^n(z) - Q_{\bar{a}_n}^n(z) \right| \log \left( \frac{Q_{\bar{a}_n}^n(z)}{Q_{\bar{b}}^n(z)} \right) \\
= \sum_{z} \left( \hat{Q}^n(z) - Q_{\bar{a}_n}^n(z) \right) \left( \log \left( \frac{Q_{\bar{a}_n}^n(z)}{Q_{\bar{b}}^n(z)} \right) \right) \mathbb{I} \left\{ Q_{\bar{a}_n}^n(z) > Q_{\bar{b}}^n(z) \right\} + \log \left( \frac{Q_{\bar{b}}^n(z)}{Q_{\bar{a}_n}^n(z)} \right) \mathbb{I} \left\{ Q_{\bar{a}_n}^n(z) < Q_{\bar{b}}^n(z) \right\} \\
\leq \sum_{z} \left( \hat{Q}^n(z) - Q_{\bar{a}_n}^n(z) \right) \left( n \log \frac{1}{v_{\min}} \mathbb{I} \left\{ Q_{\bar{a}_n}^n(z) > Q_{\bar{b}}^n(z) \right\} \\
+ \sum_{i=1}^{n} \log \frac{Q_{\bar{b}}(z_i)}{(1 - \rho_k a_n) Q_{\bar{b}}(z_i)} \mathbb{I} \left\{ Q_{\bar{a}_n}^n(z) < Q_{\bar{b}}^n(z) \right\} \\
\leq \sum_{z} \left( \hat{Q}^n(z) - Q_{\bar{a}_n}^n(z) \right) \left( n \log \frac{1}{v_{\min}} + n \log \frac{1}{(1 - \rho_k a_n)} \mathbb{I} \left\{ Q_{\bar{a}_n}^n(z) > Q_{\bar{b}}^n(z) \right\} \\
= 2 \mathcal{V}(\hat{Q}^n, Q_{\bar{a}_n}^n) \left( n \log \frac{1}{(1 - \rho_k a_n)} \mathbb{I} \left\{ Q_{\bar{a}_n}^n(z) > Q_{\bar{b}}^n(z) \right\} \\
\leq 2 \exp \left( -\frac{1}{2} \xi_{2n} a_n \right). \tag{281} \]

Defining \( v_{\min} \triangleq \min_{z} Q(a(z)) \), we bound the second term on the right hand side of (285) for \( n \) large enough as shown in (291), as shown at the top of this page, where (a) follows from the fact that \( Q_{\bar{a}_n}^n(z) \geq \prod_{k \in K} (1 - \rho_k a_n) \).
[15] D. Goeckel, B. Bash, S. Guha, and D. Towsley, “Covert communications when the warden does not know the background noise power,” IEEE Commun. Lett., vol. 20, no. 2, pp. 236–239, Feb. 2016.

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

[17] B. Bash, D. Goeckel, and D. Towsley, “Covert communication gains from adversary’s ignorance of transmission time,” IEEE Trans. Wireless Commun., vol. 15, no. 12, pp. 8394–8405, Dec. 2016.

[18] K. S. K. Arumugam and M. R. Bloch, “Keyless asynchronous covert communication,” in Proc. IEEE Int. Theory Workshop, Cambridge, U.K., Sep. 2016, pp. 191–195.

[19] K. S. K. Arumugam and M. R. Bloch, “Embedding covert information in broadcast communications,” Aug. 2018, arXiv:1808.09556. [Online]. Available: https://arxiv.org/abs/1808.09556

[20] V. Y. F. Tan and S.-H. Lee, “Time-division is optimal for covert communication over some broadcast channels,” Oct. 2017, arXiv:1710.09754. [Online]. Available: https://arxiv.org/abs/1710.09754

[21] K. S. K. Arumugam, M. R. Bloch, and L. Wang, “Covert communication over a physically degraded relay channel with non-colluding wardens,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Vail, CO, USA, Jun. 2018, pp. 766–770.

[22] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, “Covert communication achieved by a greedy relay in wireless networks,” IEEE Trans. Wireless Commun., vol. 17, no. 7, pp. 4766–4779, Jul. 2018.

[23] R. Soltani, D. Goeckel, D. Towsley, and A. Houmansadr, “Covert communications on poisson packet channels,” in Proc. Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, Sep. 2015, pp. 1046–1052.

[24] P. Mukherjee and S. Ulukus, “Covert bits through queues,” in Proc. 3rd Workshop Phys.-Layer Methods Wireless Secur., Philadelphia, PA, USA, Oct. 2016, pp. 626–630.

[25] R. Soltani, D. Goeckel, D. Towsley, B. A. Bash, and S. Guha, “Covert wireless communication with artificial noise generation,” IEEE Trans. Wireless Commun., vol. 17, no. 11, pp. 7252–7267, Nov. 2018.

[26] M. Tahmasbi and M. R. Bloch, “Covert secret key generation,” in Proc. IEEE Conf. Commun. Netw. Secur. Workshop Phys.-Layer Methods Wireless Secur., Las Vegas, NV, USA, Oct. 2017, pp. 540–544.

[27] G. Frèche, M. R. Bloch, and M. Barret, “Polar codes for covert communications over asynchronous discrete memoryless channels,” Entropy, vol. 20, no. 1, p. 3, Dec. 2017.

[28] Q. Zhang, M. Bakshi, and S. Jaggi, “Computationally efficient deniable communication,” in Proc. IEEE Int. Symp. Inf. Theory, Barcelona, Spain, Jul. 2016, pp. 2234–2238.

[29] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, “Multilevel-coded pulse-position modulation for covert communications over binary-input discrete memoryless channels,” Nov. 2018, arXiv:1811.09695. [Online]. Available: https://arxiv.org/abs/1811.09695

[30] J. Hou and G. Kramer, “Effective secrecy: Reliability, confusion and stealth,” in Proc. IEEE Int. Symp. Inf. Theory, Honolulu, HI, USA, Jun. 2014, pp. 601–605.

[31] T. S. Han and S. Verdú, “Approximation theory of output statistics,” IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 752–772, May 1993.

[32] P. Cuff, “Distributed channel synthesis,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[33] Y. Steinberg, “Resolvability theory for the multiple-access channel,” IEEE Trans. Inf. Theory, vol. 44, no. 2, pp. 472–487, Mar. 1998.

[34] M. H. Yassaee and M. R. Aref, “Multiple access wiretap channels with strong secrecy,” in Proc. IEEE Int. Theory Workshop, Dublin, Ireland, Sep. 2010, pp. 1–5.

[35] M. Frey, I. Bjelaković, and S. Stańczak, ”MAC resolvability: First and second order results,” in Proc. IEEE Conf. Commun. Netw. Secur. (CNS), Las Vegas, NV, USA, Oct. 2017, pp. 560–564.

[36] S. Verdú, “On channel capacity per unit cost,” IEEE Trans. Inf. Theory, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[37] T. M. Cover and J. A. Thomas, Elements of Information Theory. Hoboken, NJ, USA: Wiley, 2012.

[38] E. L. Lehmann and J. P. Romano, Testing Statistical Hypotheses. Springer, 2006.

[39] K. S. K. Arumugam and M. R. Bloch, “Covert communication over a K-user multiple access channel,” 2019, arXiv:1803.06007. [Online]. Available: https://arxiv.org/abs/1803.06007

[40] M. Hayashi, “General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel,” IEEE Trans. Inf. Theory, vol. 52, no. 4, Apr. 2006.

[41] T. S. Han, H. Endo, and M. Sasaki, “Reliability and secrecy functions of the wiretap channel under cost constraint,” IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 6819–6843, Nov. 2014.

[42] M. B. Parizi, E. Telatar, and N. Merhav, “Exact random coding secrecy exponents for the wiretap channel,” IEEE Trans. Inf. Theory, vol. 63, no. 1, pp. 509–531, Jan. 2017.

[43] N. Helal, M. Bloch, and A. Nosratinia, “Multiple-access channel resolvability with cribbing,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Vail, CO, USA, Jun. 2018, pp. 2052–2056.

[44] N. Helal, M. Bloch, and A. Nosratinia, “Cooperative resolvability and secrecy in the cribbing multiple-access channel,” IEEE Trans. Inf. Theory, to be published.