Covert Capacity-Key Tradeoff over Discrete Memoryless Networks

Abdelaziz Bounhar, Mireille Sarkiss, Michèle Wigger

Abstract

This paper characterizes the covert capacity-secret key tradeoffs of discrete memoryless channels (DMC) and discrete memoryless multiple-access channels (DMMAC). The focus is on channels where communication rates are measured as the number of transmitted bits divided by the square-root of the blocklength and by the square-root of the covertness constraint. Previous results had determined the largest achievable covert rates for these channels, as well as the minimum key rates required to achieve these largest covert rates. In this work, we additionally determine the minimum key rates required at reduced covert rates. Stated differently, we determine the set of all covert rates that are achievable for a given set of key rates. This is termed the *covert capacity-key tradeoff*.

Our results on the covert capacity-key tradeoffs over DMCs and DMMACs show that for small key rates and when the adversary observes the inputs through a better channel than the legitimate receiver, binary signalling at all transmitters is optimal even for larger input alphabets and the capacity-key tradeoff grows linearly. Moreover, in this regime the covert capacity-key tradeoff of DMMACs is square implying that each of the transmitters can simultaneously achieve its own largest covert rate depending only on its own available key rate, irrespective of the other transmitter. For larger key rates or when the adversary is not uniformly stronger than the legitimate receiver over all inputs, the covert capacity-key tradeoff region of DMMACs is non-square and a tradeoff arises between the rates of the various transmitters.

Index Terms

Covert communication, capacity-key tradeoff, discrete memoryless channel, multiple-access channels.

I. Introduction

As cyber threats continue to evolve, research into advanced mechanisms for securing communication links remains essential. One important area of secure communication is physical layer security. While traditional cryptographic approaches rely on computational complexity to protect information, physical layer security leverages the inherent noise and distortions in communication channels to prevent adversaries from intercepting or decoding transmissions.

An interesting subfield of physical-layer security is *covert communication*, where the goal is not to protect the content of a message but to ensure that the very act of transmission remains undetectable. This is crucial in cybersecurity, military applications, and other secure communication scenarios, where traditional encryption may still expose the presence of non-authorized communication. In covert communication instead, sophisticated coding techniques and sparse codewords are employed to ensure that the transmission is hidden in the noise of the channel, rendering detection of the transmission difficult for any adversary.

Covert communication is also relevant for securing communications in the Internet of Things (IoT) and wireless sensor networks. Indeed, many IoT applications are subject to stringent energy-limitations and can only use sparse codewords. This exactly the operating regime of covert communication, making this latter a natural candidate for securing communication in the IoT.

Though the maximum number of bits that can be sent reliably and covertly over a channel is inherently limited because codewords need to be sparse, determining the maximum amount of data bits that can be sent in covert communication is still an important question. The first characterization of this maximum number of data bits was presented for Gaussian channels in [1]. Specifically, in this work the covertness

This work was presented in part at the 2025 IEEE Information Theory Workshop. A. Bounhar was with Telecom Paris and is now with MBZUAI, Paris, France. Email: abdelaziz.bounhar@mbzuai.ac.ae. Mireille Sarkiss is with SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France. Email: mireille.sarkiss@telecom-sudparis.eu. Michèle Wigger is with LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France, Email: michele.wigger@telecom-paris.fr.

constraint was formalized by requiring that the warden's output distribution under a given coding scheme is close in Kullback-Leibler (KL) divergence to the output distribution in the absence of communication. Under these assumptions, [1] showed that the number of reliably and covertly transmitted bits cannot scale faster than the square root of the number of channel uses times the covertness constraint. This contrasts conventional communication, where transmission rates typically scale linearly with the number of channel uses. The square-root law has since been established as a fundamental principle for covert communication across a variety of channel models, including general Gaussian channels and discrete memoryless channels (DMCs) [2]–[5]. Beyond the square-root law, these results also determine the highest square-root prefactor that can be achieved while ensuring reliable and covert communication. This prefactor is often called the *covert capacity* of the channel. Notice that rates beyond the square-root regime are possible when the warden has uncertainty about the channel statistics [6]–[9], in the presence of a jammer [10]–[12], or with entanglement-assistance in quantum channels [13]–[16].

Covert communication has also been studied in networked environments with multiple transmitters and receivers. For example, [17], [18] determined the covert capacities of discrete multi-access channels (MAC) and interference channels (IC) where multiple transmitters communicate with a one or multiple receivers and where communication needs to remain undetectable to an external warden.F urther extensions have explored hybrid models that integrate covert and non-covert transmissions, revealing how the presence of non-covert users impacts covert communication rates [19]. And finally, [20], [21] examined broadcast channels (BC), where a transmitter sends a common message to all receivers while embedding a covert message for only one of them. In this scenario the attacker is thus at the same time also a legitimate receiver for certain messages.

In most of the mentioned results, covert capacities can only be achieved if transmitter(s) and legitimate receivers share a secret key when the adversary has a better channel than the legitimate receiver. Indeed, [3], [17], [18] determined the minimum key rates (measured again as the number of key bits divided by the square-root of the blocklength and the covertness constraint) required to achieve the largest possible covert rates. Since share secret keys are rare resources in practical systems, one is generally interested also in the following questions:

- What is the minimum key rate required to achieve a given covert rate?
- Which covert rates are achievable given a limited key budget?

In this article we answer these questions for discrete memoryless channels (DMC) and for discrete memoryless MACs (DMMACs). A priory these quantities can depend on how much the warden's output distribution can deviate (in KL sense) from the distribution that is induced in the absence of any communication. We shall show that for a large class of permissible KL divergences above questions share the same answer.

The largest covert rates for given limited key budgets has previously been treated in [22] for BCs and for binary-input DMCs in [19], where it was termed the *covert capacity-key tradeoff*. It was shown in [19] that this covert capacity-key tradeoff follows a linear growth for small key rates before saturating at the *covert capacity*, i.e., the largest achievable covert rate under an unlimited key budget. The results in [22] even suggested that such a linear growth also ho a general linear growth.

Our Contributions: In this paper, we characterize the covert capacity-key tradeoff over DMCs with arbitrary input alphabets. Our results illustrate that the covert capacity-key tradeoff for non-binary input DMCs critically depends on whether the KL divergence between the warden's output distributions for a given input symbol and the output distribution induced by the zero-symbol is larger or smaller than the corresponding KL divergence for the legitimate receivers. In case the warden observes a better channel than the receiver *uniformly for all non-zero inputs*, the covert capacity-key tradeoff grows linearly for low key rates and (even when the input alphabet is non-binary) the optimal signaling strategy is binary using only the zero input symbol and a specific non-zero symbol. For larger key rates the covert capacity-key tradeoff is however sublinear before it saturates at the covert capacity. In the other extreme case where the warden is uniformly (over all inputs) weaker (in above mentioned KL sense) than the legitimate receiver, the covert capacity-key tradeoff is trivially independent of the key rate. For all intermediate scenarios where the warden

is stronger than the legitimate receiver for some non-zero inputs but weaker for others, the covert-capacity key tradeoff is positive already for zero key-rates, grows sublinearly in the regime of low key rates and then saturates at the covert capacity when the key-rates are sufficiently large. In these scenarios, binary signaling is generally not optimal, not even at small key rates.

We further explore the covert capacity-key tradeoff region for the discrete memoryless multiple-access channel (DMMAC). As a first result we determine the covert capacity-key tradeoff region, i.e., the set of all rate pairs given individual key rates, for the two-user DMMAC with general finite alphabets. This tradeoff region had not been studied before and is introduced in this paper.

Similarly to the DMC scenario, the behaviour of the covert capacity-key tradeoff region of DMMACs depends on whether or not the warden is stronger than the legitimate receiver (in above described KL divergence sense). When the warden consistently observes a stronger channel across all relevant input pairs, we find that simple binary signaling is optimal (regardless of the input alphabet) for small key rates. In this regime, there is no tradeoff between the achievable covert rates of the two userseach users covert rate grows linearly with its own key rate, mirroring the behavior observed in single-user discrete memoryless channels (DMCs). However, as key rates increase, the tradeoff region grows sublinearly, and an interdependence between the covert rates of the two users emerges.

In scenarios where the warden is stronger for some input pairs but weaker for others, the covert capacity-key tradeoff region exhibits sublinear growth at all key rates and consistently presents a tradeoff between the two users achievable covert rates. Interestingly, in a special case where the warden is stronger than the legitimate receiver for all nonzero inputs from one transmitter (and for the zero input from the other) but weaker for the reverse scenario, we find that the covert capacity-key tradeoff depends only on the key rate of the first transmitter. Surprisingly, the second (but not the first) transmitter's maximum covert rate remains constant, regardless of key rate availability.

Notation: In this paper, we follow standard information theory notations. We use calligraphic fonts for sets (e.g. $\mathcal{S}$) and note by $|\mathcal{S}|$ the cardinality of a set $\mathcal{S}$. Random variables are denoted by upper case letters (e.g., X), while their realizations are denoted by lowercase letters (e.g. x). We write X^n and x^n for the tuples $(X_1, \ldots, X_n)$ and $(x_1, \ldots, x_n)$, respectively, for any positive integer $n > 0$. For a distribution P on $\mathcal{X}$, we note its product distribution on $\mathcal{X}^n$ by $P^{\otimes n}(x^n) = \prod_{i=1}^n P(x_i)$. We also denoteby $\mathrm{Supp}(P)$ the support of a distribution P, i.e. $\mathrm{Supp}(P) = \{x\colon P(x) \neq 0\}$. For two distributions P and Q on $\mathcal{X}$, $\mathbb{D}(P\|Q) = \sum_{x\in\mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$ denotes the KL divergence between the distributions. We use $\mathbb{H}(\cdot)$, $\mathbb{H}(\cdot|\cdot)$, and $\mathbb{H}_\mathrm{b}(\cdot)$ to denote entropy, conditional entropy, and binary entropy, and $\mathbb{I}(\cdot,\cdot)$ and $\mathbb{I}(\cdot;\cdot|\cdot)$ for mutual information. The logarithm function is in base 2 and motivated by continuity of the function $t\log t$ we define $0\log(0) = 0$. We use Landau notation, i.e., for a function $f(n)$ we write $f(n) = o(g(n))$ if the ratio $f(n)/g(n)$ vanishes as $n \to \infty$, and we write $f(n) = \mathcal{O}(g(n))$ if the cumulation points of the ratio $f(n)/g(n)$ are within a bounded interval. We abbreviate *probability mass function* by *pmf* and *independent and identically distributed* by *i.i.d.*.

Paper Outline: In the following Section II, we introduce covert communication over a single-user DMC and define the notion of covert capacity-key tradeoff, which is then derived and discussed in the following Section III. Section IV introduces the two-user DMMAC setup, defines the notion of covert capacity-key tradeoff region, and finally characterizes and discusses the region. Section V and the following technical appendices conclude the manuscript.

II. The Single-User Setup

Consider the setup illustrated in Figure 1. The transmitter (Tx) wishes to send a message W to the legitimate receiver (Rx) while avoiding detection by the warden which attempts to detect the presence of communication. Communication takes place over a block of n channel uses. The Tx produces channel inputs in a finite alphabet $\mathcal{X}$ and the legitimate Rx and the warden observe channel outputs within finite alphabets $\mathcal{Y}$ and $\mathcal{Z}$. These outputs are produced by a DMC, that means, if the Tx produces the n channel inputs $X^n = x^n$ then for any $i \in \{1, \ldots, n\}$ the i-th output symbols Y_i and Z_i observed at the legitimate Rx and the warden are generated from the i-th input x_i according to the conditional laws $\Gamma(\cdot|x_i)$ and $Q(\cdot|x_i)$,

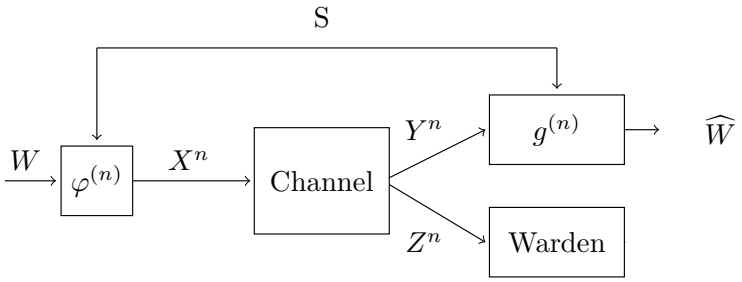

Fig. 1: Point-to-point covert communication over a Discrete Memoryless Channel.

repectively. (Notice that only the marginal conditional channel laws will matter in the sequel but not the joint conditional channel law of both outputs given the input.)

The Tx encodes message W using some encoding function $\varphi^{(n)}$ defined on appropriate domains, along with a secret-key S. Subsequently, it sends the resulting codeword

$$X^n = \varphi^{(n)}(W, S) \tag{1}$$

over the channel. For readability, we will write $x^n(w, s)$ instead of $\varphi^{(n)}(w, s)$. Let the message W and the secret-key S be represented by two sequences of m and p i.i.d. Bernoulli-$1/2$ bits, where these numbers m and p will depend on the blocklength n. The secret-key S is exclusively known to the transmitter and the Rx but not to the warden.

The legitimate Rx estimates the message as:

$$\widehat{W} = g^{(n)}(Y^n, S) \tag{2}$$

using an appropriate decoding function.

To ensure reliability of communication, we seek for systems (encoding and decoding functions) where

$$\lim_{n\to\infty} \Pr[\widehat{W} \neq W] = 0. \tag{3}$$

At the same time we impose that the output distribution implied at the warden

$$\widehat{Q}^n(z^n) \triangleq \frac{1}{2^m 2^p} \sum_{(w,s)} Q^{\otimes n}(z^n|x^n(w, s)). \tag{4}$$

be almost indistinguishable from the warden's output distribution when the all-zero sequence is transmitted (which stands for absence of communication), i.e., from

$$Q^{\otimes n}(z^n|0^n), \tag{5}$$

where we impose that the 0-symbol be part of the input alphabet $\mathcal{X}$. In particular, we will phrase the covertness constraint in terms of the KL divergence

$$D_n \triangleq \mathbb{D}\left(\widehat{Q}^n(\cdot)\middle\|Q^{\otimes n}(\cdot|0^n)\right). \tag{6}$$

A. *Covert Capacity-Key Tradeoff*

Definition 1: For any given $k \geq 0$ and sequence $\{\delta_n\}_n$, define the $\{\delta_n\}$-*capacity-key tradeoff* $r^\star(k, \{\delta_n\})$ as the largest rate $r \geq 0$ such that there exists a sequence (in the blocklength n) of tuples (m, p) and encoding/decoding functions $(\varphi^{(n)}, g^{(n)})$ satisfying

$$\lim_{n\to\infty} \Pr[\widehat{W} \neq W] = 0 \tag{7a}$$

and

$$r \leq \liminf_{n\to\infty} \frac{m}{\sqrt{n\delta_n}}, \tag{8}$$

$$k \geq \limsup_{n\to\infty} \frac{p}{\sqrt{n\delta_n}}, \tag{9}$$

and for sufficiently large blocklengths n we have $D_n \leq \delta_n$.

To ensure that the capacity-key tradeoff stays finite and is non-trivial, we assume that:

$$\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)Q(\cdot|x) \neq Q(\cdot|0), \quad \forall \psi(\cdot), \tag{10a}$$

$$\text{Supp}\left(Q(\cdot|x)\right) \subseteq \text{Supp}\left(Q(\cdot|0)\right), \quad \forall x \in \mathcal{X}, \tag{10b}$$

$$\text{Supp}\left(\Gamma(\cdot|x)\right) \subseteq \text{Supp}\left(\Gamma(\cdot|0)\right), \quad \forall x \in \mathcal{X} \tag{10c}$$

where in the above, $\psi(\cdot)$ indicates a pmf over $\mathcal{X}\backslash\{0\}$.

Moreover, we restrict to $\{\delta_n\}$ sequences that decay to 0 but no faster than in the order of $\log^2(n)/n$, i.e., we assume that:

$$\lim_{n\to\infty} \delta_n = 0, \tag{11a}$$

$$\lim_{n\to\infty} \left(\sqrt{\delta_n}\sqrt{n} - \log n\right) = \infty. \tag{11b}$$

In fact, as we show in the sequel, for all sequences $\{\delta_n\}$ satisfying (11), the capacity-key tradeoff is the same, and we shall therefore omit the argument $\{\delta_n\}$ from our notations and simply write $r^\star(k)$. Accordingly, we write the *covert capacity*, i.e., the largest value of the capacity-key tradeoff as

$$C_{\text{covert}} := \sup_{k\geq 0} r^\star(k). \tag{12}$$

Remark 1: Throughout the manuscript, our achievability results are established for sequences $\{\delta_n\}$ satisfying (11) while our converse results hold for all vanishing $\{\delta_n\}$ sequences.

III. Single-User Results

Given a pmf $\psi(\cdot)$ over $\mathcal{X}\backslash\{0\}$, we use the abbreviations

$$\chi^2(\psi) \triangleq \sum_{z\in\mathcal{Z}} \frac{\left(\sum_{x\in\mathcal{X}} \psi(x)Q(z|x) - Q(z|0)\right)^2}{Q(z|0)}. \tag{13}$$

For given $x \in \mathcal{X}\backslash\{0\}$, we define:

$$\mathbb{D}_{\mathsf{Y}}(x) \triangleq \mathbb{D}\left(\Gamma(\cdot|x)\|\Gamma(\cdot|0)\right), \tag{14}$$

$$\mathbb{D}_{\mathsf{Z}}(x) \triangleq \mathbb{D}\left(Q(\cdot|x)\|Q(\cdot|0)\right). \tag{15}$$

Finally, for each pmf ψ over $\mathcal{X}\backslash\{0\}$, we define a function $r_\psi\colon k \mapsto r_\psi(k)$. If the pmf ψ is such that the difference

$$\Delta_\psi \triangleq \sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot\left(\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x)\right) \tag{16}$$

is positive, we define

$$r_\psi(k) = \min\left\{\frac{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot\mathbb{D}_{\mathsf{Y}}(x)}{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot\left(\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x)\right)}\cdot k,\ \sqrt{2}\frac{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot\mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi)}}\right\}, \tag{17}$$

and if ψ is such that $\Delta_\psi \leq 0$, we define

$$r_\psi(k) = C_\psi \triangleq \sqrt{2}\frac{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot\mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi)}}. \tag{18}$$

So, if $\Delta_\psi \leq 0$ the function is constant equal to C_ψ and if $\Delta_\psi > 0$, the function grows linearly from the origin $(k = 0, r = 0)$ to the point

$$k = k_\psi \triangleq \sqrt{2}\frac{\sum_{x\in\mathcal{X}\setminus\{0\}} \psi(x)\cdot(\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x))}{\sqrt{\chi^2(\psi)}} \tag{19}$$

$$r = C_\psi \triangleq \sqrt{2}\frac{\sum_{x\in\mathcal{X}\setminus\{0\}} \psi(x)\cdot\mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi)}}, \tag{20}$$

and saturates at $r_\psi(k) = C_\psi$ for all key rates $k \geq k_\psi$.

Theorem 1: We have

$$r^\star(k) = \max_\psi r_\psi(k), \tag{21}$$

where the maximum is over all pmfs ψ on $\mathcal{X}\setminus\{0\}$. Moreover, the covert capacity can be expressed as

$$C_{\text{covert}} := \sup_\psi C_\psi. \tag{22}$$

Proof: Can formally be obtained by specializing our MAC result, Theorem 5 ahead, to the case where one of the input alphabets is degenerate. We provide here a brief sketch of a direct proof for the DMC.

Sketch of Achievability: Fix ψ and parameter

$$\phi = \min\left\{\frac{k^2}{k_\psi^2}, 1\right\}. \tag{23}$$

Apply the coding scheme in [3, Section V] to only the first $n_1 = \lfloor \phi n \rfloor$ channel uses and transmit the zero-symbol otherwise. Probability of error can tend to 0 for

$$m \approx \phi\omega_n\sqrt{n}\sum_{x\in\mathcal{X}\setminus\{0\}} \psi(x)\mathbb{D}_{\mathsf{Y}}(x), \tag{24a}$$

$$m + p \approx \phi\omega_n\sqrt{n}\sum_{x\in\mathcal{X}\setminus\{0\}} \psi(x)\mathbb{D}_{\mathsf{Z}}(x), \tag{24b}$$

and with KL-divergence

$$D_n \approx \phi\cdot\frac{\omega_n^2}{2}\chi^2(\psi). \tag{25}$$

for any sequence ω_n tending to 0 slowlier than $\log^2(n)/n$. Combining these results proves achievability of the rate $r_\psi(k)$.

Sketch of Converse: Let k be fixed and consider any sequence of encodings and decodings satisfying (7) and $D_n \leq \delta_n$ for sufficiently large n. By the considerations in [3], there must exist a sequence ω_n tending to 0 and a pmf ψ over $\mathcal{X}\setminus\{0\}$ so that (24) and (25) hold with inequalities $\leq$, $\geq$, $\geq$, respectively, and for ϕ arbitrary close to 1 (above or belwo depending on the direction of the inequality). Since $\delta_n \geq D_n$, for sufficiently large values of n, we can conclude that there must exist a parameter $\beta \in [0, 1]$ so that $\delta_n \approx \beta^{-2}\cdot\frac{\omega_n^2}{2}\chi^2(\psi)$. Solving for ω_n and plugging into the mentioned inequalities corresponding to (24) roughly proves that

$$r \leq \sqrt{2}\beta\frac{\sum_{x\in\mathcal{X}\setminus\{0\}} \psi(x)\cdot\mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi)}} \tag{26}$$

$$k \geq \sqrt{2}\beta\frac{\sum_{x\in\mathcal{X}\setminus\{0\}} \psi(x)\cdot(\mathbb{D}_{\mathsf{Z}(x)-\mathbb{D}_{\mathsf{Y}}}(x))}{\sqrt{\chi^2(\psi)}}, \tag{27}$$

for some pmf $\psi(\cdot)$ and some $\beta \in [0, 1]$. Optimizing finally over β and ψ proves that $r^*(k) \leq \sup_\psi r_\psi(k)$. ■

Remark 2: Notice that for binary input-channels there is only one valid distribution ψ (the singleton on the non-zero input) and Theorem 1 simplifies considerably.

Remark 3: As follows from the proof of Theorem 1, $r_\psi(k)$ is the largest covert rate that can be achieved with a given input pmf ψ and key rate k.

The approach proposed in [22] of appropriately choosing the ω_n-sequence in [3] (see also our proof sketch above) does not suffice to determine the optimal set of message- and key-rates that can be achieved. In fact, scaling the ω_n-sequence does not impact the rates r and k. Notice further that for non-binary input alphabets the dependency of the message- and key-rates, i.e., our capacity-key tradeoff, is not necessarily linear as claimed in [22]. The following subsection discusses regimes where $r^\star(k)$ is linear and where it is not.

A. Discussion and Simplifications of $r^\star(k)$

For the further analysis of $r^\star(k)$, we distinguish three cases depending on whether Δ_ψ is positive for all pmfs ψ, is negative for all ψ, or it is positive for some ψs and negative for others. Consider an example where Δ_ψ is always positive:

Example 1 (Adversary Stronger than Legitimate Receiver): Consider a channel with input alphabet $\mathcal{X} = \{0,1,2\}$, output alphabets $\mathcal{Z} = \mathcal{Y} = \{0,1,2,3,4\}$, and transition pmfs:

$$\Gamma = \begin{bmatrix} 0.23, & 0.20, & 0.25, & 0.05, & 0.27 \\ 0.35, & 0.22, & 0.10, & 0.05, & 0.28 \\ 0.27, & 0.24, & 0.24, & 0.07, & 0.18 \end{bmatrix} \tag{28a}$$

and

$$Q = \begin{bmatrix} 0.22, & 0.32, & 0.15, & 0.12, & 0.19 \\ 0.36, & 0.03, & 0.39, & 0.21, & 0.01 \\ 0.31, & 0.20, & 0.07, & 0.22, & 0.20 \end{bmatrix} \tag{28b}$$

where the rows pertain to the three input symbols $x = 0, 1, 2$ and the columns to the five output symbols $0, 1, 2, 3, 4$.

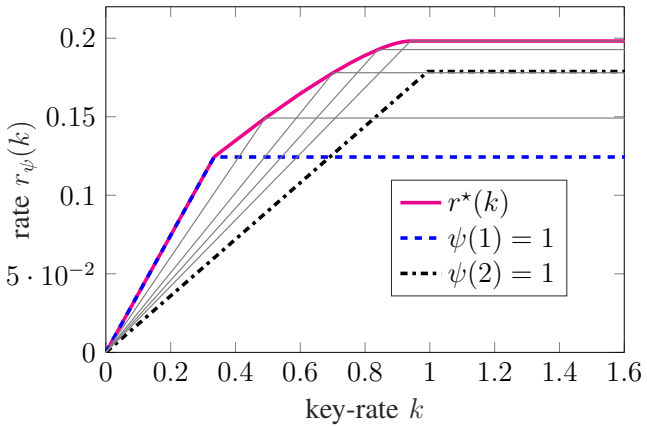

Fig. 2: The figure illustrates the functions $r_\psi(k)$ for different pmfs ψ. The capacity-secret-key tradeoff $r^\star(k)$ corresponds to the upper convex hull of all the curves.

For this channel the difference $\Delta_\psi > 0$ for all pmfs ψ, since for all $x \in \mathcal{X}\backslash\{0\}$ we have $\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x) > 0$. Figure 2 illustrates the function r_ψ for different choices of the input pmf ψ. The blue dashed line corresponds to the degenerate choice $\psi(1) = 1$ and the black dash-dotted line to the degenerate choice $\psi(2) = 1$.

We observe that in this example the extreme lines $r_\psi(k)$ with smallest and largest slopes correspond to the degenerate pmfs that put probability 1 on a single non-zero input. This is true in general as can be seen by the following lemma and noting that the slope of $r_\psi(k)$ is given by

$$S_\psi \triangleq \frac{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot \mathbb{D}_{\mathsf{Y}}(x)}{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot (\mathbb{D}_{\mathsf{Z}}(x)-\mathbb{D}_{\mathsf{Y}}(x))} > 0. \tag{29}$$

Lemma 1: If

$$\mathbb{D}_{\mathsf{Z}}(x)-\mathbb{D}_{\mathsf{Y}}(x) > 0, \qquad \forall x\in\mathcal{X}\backslash\{0\}, \tag{30}$$

then $S_\psi > 0$ for all input pmfs ψ and is largest (smallest) for a degenerate pmf ψ that puts probability mass 1 on one of the non-zero inputs.

Proof: Consider the convex hull of the points

$$\{(\mathbb{D}_{\mathsf{Z}}(x)-\mathbb{D}_{\mathsf{Y}}(x),\ \mathbb{D}_{\mathsf{Y}}(x))\}_{x\in\mathcal{X}\backslash\{0\}} \tag{31}$$

which by our assumption all lie in the first quadrant of the two-dimensional Euclidean space. If for each point we consider the straight line from the origin to this point, the line with largest (smallest) slope is attained at one of the points in (31). Moreover, the slope for each of the points in (31) is given by

$$\frac{\mathbb{D}_{\mathsf{Y}}(x)}{\mathbb{D}_{\mathsf{Z}}(x)-\mathbb{D}_{\mathsf{Y}}(x)}, \tag{32}$$

and thus equal to S_ψ for degenerate pmf ψ with probability mass 1 on the non-zero symbol x.

In a similar way, since each point $P=(P_1,P_2)$ in the convex hull is naturally assigned to a different pmf ψ in the sense that:

$$P_1 = \sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot(\mathbb{D}_{\mathsf{Z}}(x)-\mathbb{D}_{\mathsf{Y}}(x)) \tag{33}$$

$$P_2 = \sum_{x\in\mathcal{X}\backslash\{0\}} \psi(x)\cdot\mathbb{D}_{\mathsf{Y}}(x), \tag{34}$$

and the slope of the straight line from the origin to this P is given by S_ψ, from the arguments in the previous paragraphs, we can deduce that the largest (smallest) slope S_ψ is obtained from a degenerate pmf ψ with probability mass on a single non-zero symbol. ■

With above lemma, we can simplify Theorem 1 for the class of channels satisfying (30).

Corollary 2 (Adversary Stronger than Legitimate Receiver): If (30) holds, then

$$r^\star(k) = \begin{cases} S_{\max}\cdot k & \text{if } \in [0,\ k_{\text{lin}}]\,, \\ \max_{\psi\in\mathcal{L}(k)} C_\psi & \text{if } \in (k_{\text{lin}},\ k_{\text{sat}})\,, \\ C_{\text{covert}} & \text{if } \in [k_{\text{sat}},\infty)\,, \end{cases} \tag{35}$$

where

$$S_{\max} \triangleq \max_{x\in\mathcal{X}\backslash\{0\}} \frac{\mathbb{D}_{\mathsf{Y}}(x)}{\mathbb{D}_{\mathsf{Z}}(x)-\mathbb{D}_{\mathsf{Y}}(x)}; \tag{36}$$

and

$$k_{\text{lin}} \triangleq \sqrt{2}\frac{\mathbb{D}_{\mathsf{Z}}(x_{\text{best}})-\mathbb{D}_{\mathsf{Y}}(x_{\text{best}})}{\sqrt{\chi^2(\delta_{x_{\text{best}}})}} \tag{37}$$

$$k_{\text{sat}} \triangleq \sqrt{2}\frac{\sum_{x\in\mathcal{X}\backslash\{0\}} \psi^*(x)\cdot\mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi^*)}}, \tag{38}$$

where x_{best} is the maximizer in (36); $\delta_{x_{\text{best}}}$ indicates the degenerate pmf with probability 1 at x_{best}; and ψ^* the maximizer in (22), i.e., the covert capacity achieving input-distribution. Moreover,

$$\mathcal{L}(k) \triangleq \left\{ \psi \colon \quad k \geq \sqrt{2} \frac{\sum_{x \in \mathcal{X}\backslash\{0\}} \psi(x) \cdot \mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi)}} \right\}. \tag{39}$$

Remark 4: From Lemma 1 and above Corollary 2, we can directly deduce that for channels satisfying (30), binary signaling on inputs 0 and x_{best} is optimal for small key rates $k \leq S_{\max}$, where recall that x_{best} is the maximizing input in (36) and $S_{\max}$ is defined in (36). Moreover, in the regime of small key rates the covert capacity-key tradeoff is linear. Specifically, increasing the key-rate by 1 will increase the largest achievable rate by $S_{\max}$.

For channels where (30) holds, the capacity-key tradeoff starts at the origin: $r^\star(0) = 0$. For all other channels however this is not the case. In fact, when the adversary is no better than the legitimate Rx for at least one of the non-zero inputs, i.e., when

$$\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x) \leq 0, \qquad \text{for some } x \in \mathcal{X}\backslash\{0\}, \tag{40}$$

then a positive message-rate is achievable even with zero key-rate, i.e., $r^\star(0) > 0$. In fact, we have the following corollary directly from Corollary 1 and our discussion above.

Corollary 3 (Adversary Sometimes Weaker than Legitimate Rx): If (40) holds, then

$$r^\star(k) = \max_{\psi \in \mathcal{L}(k)} \sqrt{2} \frac{\sum_{x \in \mathcal{X}\backslash\{0\}} \psi(x) \cdot \mathbb{D}_{\mathsf{Y}}(x)}{\sqrt{\chi^2(\psi)}}. \tag{41}$$

In particular,

$$r^\star(k) = C_{\text{covert}}, \qquad k \geq k_{\text{sat}}. \tag{42}$$

The following example illustrates the covert capacity tradeoff for a channel satisfying Assumption (40).

Example 2 (Adversary Sometimes Weaker than Legitimate Rx): Consider a channel with ternary inputs $\mathcal{X} = \{0, 1, 2\}$, quinary outputs $\mathcal{Y} = \mathcal{Z} = \{0, 1, 2, 3, 4\}$, and transition pmfs:

$$\Gamma = \begin{bmatrix} 0.24 & 0.10 & 0.22 & 0.22 & 0.22 \\ 0.20 & 0.14 & 0.26 & 0.328 & 0.072 \\ 0.06 & 0.19 & 0.2 & 0.05 & 0.50 \end{bmatrix} \tag{43a}$$

and

$$Q = \begin{bmatrix} 0.32 & 0.22 & 0.23 & 0.13 & 0.10 \\ 0.47 & 0.25 & 0.10 & 0.14 & 0.04 \\ 0.38 & 0.01 & 0.15 & 0.12 & 0.34 \end{bmatrix} \tag{43b}$$

For this channel, Assumption (40) holds for $x = 1$ but not for $x = 2$.

Figure 3 illustrates the function r_ψ for above channel and for different choices of the input pmf ψ. The blue dashed line corresponds to the degenerate choices $\psi(1) = 1$ and the black dash-dotted line to $\psi(2) = 1$.

To study the last case where

$$\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x) \leq 0, \qquad \text{for all } x \in \mathcal{X}\backslash\{0\}, \tag{44}$$

we notice that this implies in particular also that for any ψ:

$$\sum_{x \in \mathcal{X}\backslash\{0\}} \psi(x) \left(\mathbb{D}_{\mathsf{Z}}(x) - \mathbb{D}_{\mathsf{Y}}(x) \right) \leq 0 \tag{45}$$

and thus each of the functions r_ψ is a straigth line.

Corollary 4 (Adversary Weaker than Legitimate Rx): Under Condition (44),

$$r^\star(k) = C_{\text{covert}}, \quad k \geq 0. \tag{46}$$

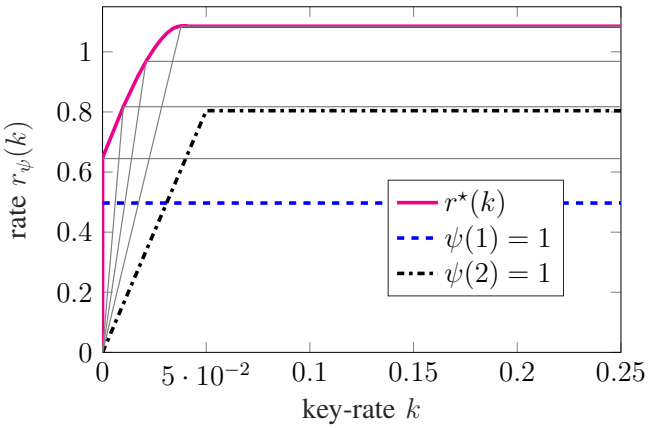

Fig. 3: The figure illustrates the functions $r_\psi(k)$ for different pmfs ψ. The capacity-secret-key tradeoff $r^\star(k)$ corresponds to the upper convex hull of all the curves.

IV. The Multiple-Access Channel: Model and Results

We turn our focus to a two-user multi-access channel (MAC), see Figure 4, where the two Txs produce inputs in finite alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, the legitimate Rx observes outputs in the finite alphabet $\mathcal{Y}$, and the warden observes outputs in the finite alphabet $\mathcal{Z}$.

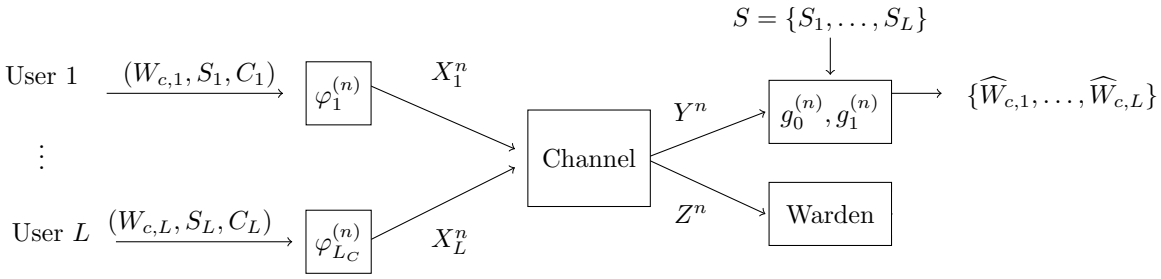

Fig. 4: Covert communication over a two-user MAC.

This covert MAC has previously been studied in [17]. The main difference here is that each Tx $j\in\{1,2\}$ has access to additional *local* randomness described by C_i which consists of g_j i.i.d. Bernoulli-$1/2$ bits. Each Tx $j\in\{1,2\}$ thus produces its channel inputs as

$$X_j^n=\varphi^{(n)}(W_j,S_j,C_j), \tag{47}$$

where W_j and S_j are independent i.i.d. Bernoulli-$1/2$ bitstrings of lengths m_j and p_j (which grow with n), respectively and S_j is know to Tx j and the Rx, while W_j and C_j only to Tx j.

As for the single-user setup, the legitimate Rx and the warden observe outputs generated by discrete memoryless channels $\Gamma(\cdot|\cdot,\cdot)$ and $Q(\cdot|\cdot,\cdot)$ based on the input sequences produced at the two Txs. That means, if Tx 1 sends inputs $X_1^n=x_1^n$ and Tx 2 sends $X_2^n=x_2^n$, then for each $i\in\{1,\dots,n\}$, the legitimate Rx observes output symbol Y_i following the conditional law $\Gamma(\cdot|x_{1,i},x_{2,i})$ and the adversary observes output Z_i following the conditional law $Q(\cdot|x_{1,i},x_{2,i})$.

After observing outputs Y^n, the legitimate Rx decodes both messages W_1 and W_2 as:

$$(\widehat{W}_1,\widehat{W}_2)=g^{(n)}(Y^n,S_1,S_2). \tag{48}$$

using an appropriate decoding function $g^{(n)}$.

$$\chi^2(\psi) \triangleq \sum_{z\in\mathcal{Z}} \frac{\left(\sum_{(x_1,x_2)\in\tilde{\mathcal{X}}} \psi(x_1,x_2)Q(z|x_1,x_2) - Q(z|0,0)\right)^2}{Q(z|0,0)}. \tag{56}$$

Covertness imposes that the output distribution implied at the warden

$$\widehat{Q}^n(z^n) \triangleq \frac{\sum_{\substack{(w_1,w_2,s_1,\\ s_2,c_1,c_2)}} Q^{\otimes n}(z^n|x_1^n(w_1,s_1,c_1),x_2^n(w_2,s_2,c_2))}{2^{m_1+m_2}2^{p_1+p_2}2^{g_1+g_2}}. \tag{49}$$

be almost indistinguishable from the warden's output distribution $Q^{\otimes n}(\cdot|0^n,0^n)$ when the all-zero sequence is transmitted by both Txs (i.e. no communication). We thus assume that both input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ contain the zero-symbol 0, and then require that the KL-divergence

$$D_n \triangleq \mathbb{D}\left(\widehat{Q}^n \middle\| Q^{\otimes n}(\cdot|0^n,0^n)\right) \tag{50}$$

be below a target sequence δ_n.

Definition 2: For a given sequence $\{\delta_n\}$ and key rates $k_1,k_2\geq 0$, define the $\{\delta_n\}$-*capacity-key tradeoff region* $\mathcal{R}^\star_{\{\delta_n\}}(k_1,k_2)$ as the set of all rate pairs (r_1,r_2) such that there exists a sequence (in the blocklength n) of tuples $(m_1,m_2,p_1,p_2,g_1,g_2)$ and encoding/decoding functions $(\varphi_1^{(n)},\varphi_2^{(n)},g^{(n)})$ satisfying

$$\lim_{n\to\infty}\Pr[(\widehat{W}_1,\widehat{W}_2)\neq(W_1,W_2)]=0, \tag{51}$$

and

$$r_j \leq \liminf_{n\to\infty}\frac{m_j}{\sqrt{n\delta_n}}, \quad j\in\{1,2\}, \tag{52}$$

$$k_j \geq \limsup_{n\to\infty}\frac{p_j}{\sqrt{n\delta_n}}, \quad j\in\{1,2\}, \tag{53}$$

and for sufficiently large blocklengths n we have $D_n\leq\delta_n$.

Again, we shall show that for all sequences $\{\delta_n\}$ satisfying (11), the capacity-key tradeoff region is the same, and therefore in the sequel we simply write $\mathcal{R}^\star(k_1,k_2)$.

To avoid that the problem be trivial or impossible, we impose the following restrictions on the MAC to hold for any input pair $(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2$:

$$\sum_{\psi}\psi(x_1,x_2)Q(\cdot|x_1,x_2)\neq Q(\cdot|0,0), \quad \forall\psi, \tag{54a}$$

$$\text{Supp}(\Gamma(\cdot|x_1,x_2))\subseteq\text{Supp}(\Gamma(\cdot|0,0)), \quad \forall x_1\in\mathcal{X}_1,x_2\in\mathcal{X}_2 \tag{54b}$$

$$\text{Supp}(Q(\cdot|x_1,x_2))\subseteq\text{Supp}(Q(\cdot|0,0)), \quad \forall x_1\in\mathcal{X}_1,x_2\in\mathcal{X}_2. \tag{54c}$$

where here ψ denotes any pmf over $\mathcal{X}_1\times\mathcal{X}_2$.

A. Results

Define the following subset of $\mathcal{X}_1\times\mathcal{X}_2$:

$$\tilde{\mathcal{X}} \triangleq ((\mathcal{X}_1\backslash\{0\}) \times \{0\}) \cup (\{0\}\times(\mathcal{X}_2\backslash\{0\})). \tag{55}$$

For any pmf ψ over $\tilde{\mathcal{X}}$, define the quantity (56) on top of the page. Define further:

$$\mathbb{D}_\Upsilon(x_1,x_2) = \mathbb{D}\left(\Gamma(\cdot|x_1,x_2)\,\|\,\Gamma(\cdot|0,0)\right), \tag{57}$$

$$r_1 \leq \min\left\{ \frac{\sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1,0)\cdot \mathbb{D}_{\mathsf{Y}}(x_1,0)\cdot k_1}{\max\left\{\sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1,0)\cdot\left(\mathbb{D}_{\mathsf{Z}}(x_1,0)-\mathbb{D}_{\mathsf{Y}}(x_1,0)\right),\ 0\right\}},\ \sqrt{2}\frac{\sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1,0)\cdot \mathbb{D}_{\mathsf{Y}}(x_1,0)}{\sqrt{\chi^2(\psi)}} \right\} \tag{59a}$$

$$r_2 \leq \min\left\{ \frac{\sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0,x_2)\cdot \mathbb{D}_{\mathsf{Y}}(0,x_2)\cdot k_2}{\max\left\{\sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0,x_2)\cdot\left(\mathbb{D}_{\mathsf{Z}}(0,x_2)-\mathbb{D}_{\mathsf{Y}}(0,x_2)\right),\ 0\right\}},\sqrt{2}\frac{\sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0,x_2)\cdot \mathbb{D}_{\mathsf{Y}}(0,x_2)}{\sqrt{\chi^2(\psi)}} \right\}. \tag{59b}$$

$$\mathbb{D}_{\mathsf{Z}}(x_1,x_2) = \mathbb{D}\left(Q(\cdot|x_1,x_2)\,\|\,Q(\cdot|0,0)\right). \tag{58}$$

For any fixed ψ and key-pairs (k_1,k_2), let $\mathcal{R}^\star_\psi(k_1,k_2)$ denote the rate-pairs (r_1,r_2) satisfying Inequalities (59) on top of the page, where we define $a/0=\infty$.

Theorem 5: Given key-rates $k_1,k_2\geq 0$ and a sequence $\{\delta_n\}_{n\geq 1}$ satisfying (11), the capacity-key tradeoff-region is

$$\mathcal{R}^\star(k_1,k_2) = \bigcup_{\psi} \mathcal{R}^\star_\psi(k_1,k_2), \tag{60}$$

where the union is taken over all pmfs ψ over the alphabet $\tilde{\mathcal{X}}$.

Proof: See Appendix A. ∎

Remark 5 (Local Randomness only Required at One Transmitter): Inspecting the achievability proof in Appendix A, we notice that given key rates k_1 and k_2 and pmf ψ over $\tilde{\mathcal{X}}$ the entire region $\mathcal{R}^\star_\psi(k_1,k_2)$ is

- achievable without local randomness at both Txs if the maxima in (59) are achieved by the second terms.
- achievable without common randomness at Tx 1 if

$$k_1 \sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0,x_2)\cdot\left(\mathbb{D}_{\mathsf{Z}}(0,x_2)-\mathbb{D}_{\mathsf{Y}}(0,x_2)\right) > k_2 \sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1,0)\cdot\left(\mathbb{D}_{\mathsf{Z}}(x_1,0)-\mathbb{D}_{\mathsf{Y}}(x_1,0)\right)$$

- achievable without common randomness at Tx 2 if

$$k_1 \sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0,x_2)\cdot\left(\mathbb{D}_{\mathsf{Z}}(0,x_2)-\mathbb{D}_{\mathsf{Y}}(0,x_2)\right) < k_2 \sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1,0)\cdot\left(\mathbb{D}_{\mathsf{Z}}(x_1,0)-\mathbb{D}_{\mathsf{Y}}(x_1,0)\right).$$

Remark 6 (Opacity to channel transition probabilities for non-zero input pairs): Examining the expressions in (59), one can notice that the covert capacity-key tradeoff only depends on the channel transition probabilities $\Gamma(\cdot|x_1,x_2)$ and $Q(\cdot|x_1,x_2)$ for input pairs $(x_1,x_2)\in\tilde{\mathcal{X}}\cup\{(0,0)\}$ but not on the transition probabilities for other input pairs.

The structure of the covert capacity-key tradeoff region $\mathcal{R}^\star(k_1,k_2)$ depends on whether the differences

$$\mathbb{D}_{\mathsf{Z}}(x_1,0)-\mathbb{D}_{\mathsf{Y}}(x_1,0)>0, \qquad x_1\in\mathcal{X}_1\backslash\{0\}, \tag{61a}$$

$$\mathbb{D}_{\mathsf{Z}}(0,x_2)-\mathbb{D}_{\mathsf{Y}}(0,x_2)>0, \qquad x_2\in\mathcal{X}_2\backslash\{0\} \tag{61b}$$

are positive or not.

Remark 7 (When no Key is Required at a Tx): When (61a) is violated for all inputs $x_1\in\mathcal{X}_1\backslash\{0\}$, then the r_1-rate only depends on the choice of the pmf ψ but not on k_1. Moreover, the extreme r_1-point of the covert capacity-key tradeoff region $\mathcal{R}^\star(k_1,k_2)$ neither depends on the key rate k_2; the region itself however does. Similar observations hold for r_2 when (61b) holds for all inputs $x_2\in\mathcal{X}_2\backslash\{0\}$. Obviously, if (61) holds for all non-zero inputs x_1 and x_2, then the entire capacity-key tradeoff region depends neither on k_1 nor on k_2.

In contrast, when the adversary is uniformly *stronger* than the legitimate Rx, the following result holds.

Corollary 6 (Adversary Stronger than Legitimate Receiver): If (61) hold for all $x_1 \in \mathcal{X}_1\backslash\{0\}$ and $x_2 \in \mathcal{X}_2\backslash\{0\}$, then

$$\mathcal{R}^\star(k_1,k_2) = \begin{cases} [0, k_1 S_{1,\max}] \times [0, k_2 S_{2,\max}] & \text{if } k_1 \le k_{1,\text{lin}} \\ & \quad k_2 \le k_{2,\text{lin}}, \\ \max\limits_{\psi \in \mathcal{L}(k_1,k_2)} \mathcal{R}^\star_\psi(k_1,k_2) & \text{if } k_1 \ge k_{1,\text{lin}}, \\ & \text{or } k_2 \ge k_{2,\text{lin}}, \end{cases} \tag{62}$$

where

$$S_{1,\max} \triangleq \max_{x_1 \in \mathcal{X}_1\backslash\{0\}} \frac{\mathbb{D}_{\mathsf{Y}}(x_1,0)}{\mathbb{D}_{\mathsf{Z}}(x_1,0) - \mathbb{D}_{\mathsf{Y}}(x_1,0)}, \tag{63a}$$

$$S_{2,\max} \triangleq \max_{x_2 \in \mathcal{X}_2\backslash\{0\}} \frac{\mathbb{D}_{\mathsf{Y}}(0,x_2)}{\mathbb{D}_{\mathsf{Z}}(0,x_2) - \mathbb{D}_{\mathsf{Y}}(0,x_2)}, \tag{63b}$$

and

$$k_{1,\text{lin}} \triangleq \sqrt{2}\frac{\mathbb{D}_{\mathsf{Z}}(x_{1,\text{best}},0) - \mathbb{D}_{\mathsf{Y}}(x_{1,\text{best}},0)}{\sqrt{\chi^2(\delta_{x_1,0})}}, \tag{64}$$

$$k_{2,\text{lin}} \triangleq \sqrt{2}\frac{\mathbb{D}_{\mathsf{Z}}(0,x_{2,\text{best}}) - \mathbb{D}_{\mathsf{Y}}(0,x_{2,\text{best}})}{\sqrt{\chi^2(\delta_{0,x_2})}}, \tag{65}$$

for $x_{1,\text{best}}$ and $x_{2,\text{best}}$ the maximizers in (63) and

$$\mathcal{L}(k_1,k_2) \triangleq \Bigg\{\psi\colon\ k_1 \ge \sqrt{2}\frac{\sum_{x_1 \in \mathcal{X}_1\backslash\{0\}} \psi(x_1,0)\cdot \mathbb{D}_{\mathsf{Y}}(x_1,0)}{\sqrt{\chi^2(\psi)}}, \\ k_2 \ge \sqrt{2}\frac{\sum_{x_2 \in \mathcal{X}_2\backslash\{0\}} \psi(0,x_2)\cdot \mathbb{D}_{\mathsf{Y}}(0,x_2)}{\sqrt{\chi^2(\psi)}}\Bigg\}. \tag{66}$$

Proof: Under Condition (61), the first term in (59) is stringent for small key values k_1, k_2 because the denominator is non-zero. Moreover, these first expressions do not depend on the overall pmf ψ, but only on the x_1- and x_2- "marginals", respectively. Moreover, the capacity-key tradeoff region is achieved by a singleton distribution putting all mass on a single non-zero value for x_1 and x_2. Details of the proofs are omitted due to lack of space. ■

Remark 8 (Optimality of Binary Signaling for Strong Adversaries and Small Key Rates): Binary signaling (i.e. $\mathcal{X}_j = \{0, x_{j,\text{best}}\}, \forall j \in \{1,2\}$) at both Txs is optimal when the adversary is stronger than the legitimate Rx and in the regime of small key rates. Moreover, in this regime there is no tradeoff between the largest covert rates that the two users can simultaneously achieve. For larger key rates however a tradeoff arises between the largest rates that are simultaneously achievable at the two users.

We end this subsection with two numerical examples.

Example 3 (Ternary Inputs with a Stronger Adversary): Consider now a DMMAC with ternary input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1,2\}$ and let the output alphabets $\mathcal{Y} = \mathcal{Z} = \{0,1,2,3\}$. Assume that the channel transition laws are given by

$$\Gamma = \begin{bmatrix} 0.2 & 0.28 & 0.28 & 0.24 \\ 0.05 & 0.1 & 0.45 & 0.4 \\ 0.07 & 0.37 & 0.4 & 0.16 \\ 0.23 & 0.24 & 0.2 & 0.33 \\ 0.18 & 0.4 & 0.14 & 0.28 \\ 0.2 & 0.27 & 0.5 & 0.03 \\ 0.05 & 0.35 & 0.49 & 0.11 \\ 0.2 & 0.42 & 0.25 & 0.13 \\ 0.32 & 0.47 & 0.01 & 0.2 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 0.2 & 0.2 & 0.36 & 0.24 \\ 0.01 & 0.37 & 0.17 & 0.45 \\ 0.42 & 0.35 & 0.05 & 0.18 \\ 0.25 & 0.25 & 0.15 & 0.37 \\ 0.6 & 0.27 & 0.01 & 0.12 \\ 0.3 & 0.1 & 0.33 & 0.27 \\ 0.1 & 0.22 & 0.67 & 0.01 \\ 0.38 & 0.32 & 0.25 & 0.05 \\ 0.17 & 0.3 & 0.23 & 0.3 \end{bmatrix}, \tag{67a}$$

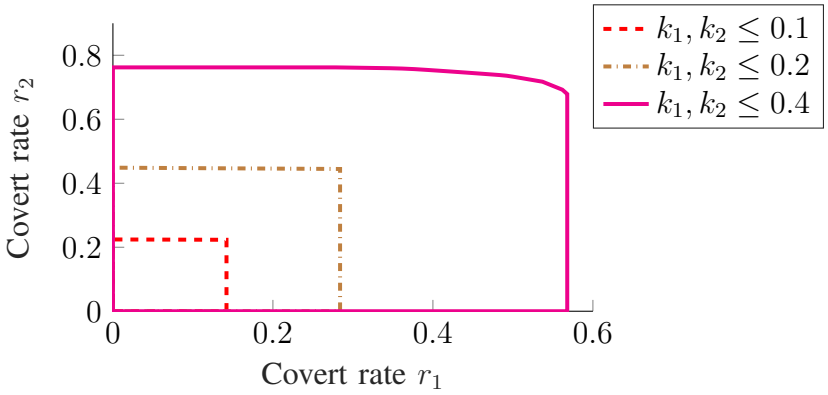

Fig. 5: $\mathcal{R}^*(k_1, k_2)$ for the channel in (67) under various symmetric constraints on the key rates k_1 and k_2.

where the four columns correspond to the four output symbols $0, 1, 2, 3$ and the nine rows correspond to the nine distinct pairs (x_1, x_2) in increasing alphabetical ordering, i.e., $(0,0), (0,1), (1,0), (1,1), \ldots, (2,2)$. According to Theorem 5, since ψ only takes value over the input pairs $(x_1, x_2) \in \tilde{\mathcal{X}}$ where exactly one of the two inputs is zero, the covert capacity-key tradeoff region only depends on rows $1, 2, 3, 4$, and 7. The other rows in both transition matrices Γ and Q do not play any role.

For this channel Conditions (61) are satisfied for all non-zero inputs x_1, x_2. That means the adversary is again uniformly (over all inputs) stronger than the legitimate Rx (in KL-sense) and thus positive key rates are required to achieve positive covert rates. Figure 5 illustrates the capacity key-rate tradeoff regions for this channel at different secret-key rates. We observe that for small key rates, the tradeoff region is a square region and there is no tradeoff between the largest covert rates r_1 and r_2 that are simultaneously achievable at the two users. For larger key rates a tradeoff arises between the two covert rates.

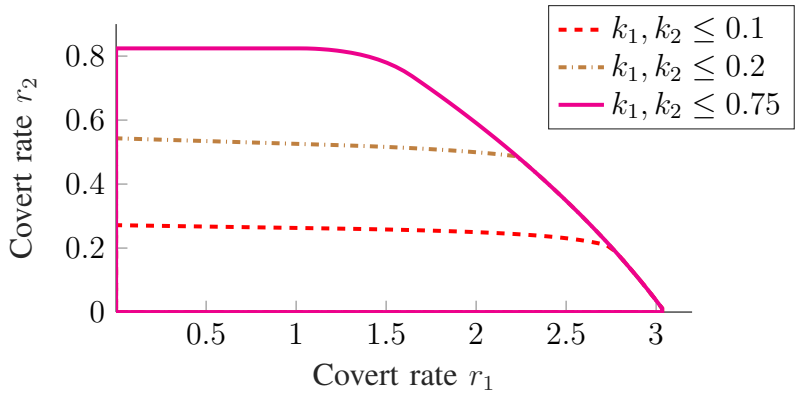

Fig. 6: $\mathcal{R}^*(k_1, k_2)$ for the channel in (68) when symmetric constraints are imposed on the key rates k_1 and k_2.

Example 4 (Ternary Example with a Mixed (Stronger/Weaker) Adversary): Consider another DMMAC with ternary input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2\}$, output alphabets $\mathcal{Y} = \mathcal{Z} = \{0, 1, 2, 3\}$, and channel

transition laws

$$\Gamma = \begin{bmatrix} 0.20 & 0.29 & 0.28 & 0.23 \\ 0.05 & 0.10 & 0.44 & 0.41 \\ 0.07 & 0.37 & 0.40 & 0.16 \\ 0.31 & 0.47 & 0.01 & 0.21 \\ 0.23 & 0.24 & 0.21 & 0.33 \\ 0.18 & 0.4 & 0.14 & 0.29 \\ 0.21 & 0.27 & 0.50 & 0.02 \\ 0.04 & 0.35 & 0.50 & 0.11 \\ 0.21 & 0.42 & 0.25 & 0.12 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 0.20 & 0.19 & 0.36 & 0.25 \\ 0.01 & 0.37 & 0.17 & 0.45 \\ 0.42 & 0.35 & 0.05 & 0.18 \\ 0.17 & 0.31 & 0.22 & 0.30 \\ 0.25 & 0.25 & 0.19 & 0.31 \\ 0.60 & 0.27 & 0.02 & 0.11 \\ 0.29 & 0.10 & 0.33 & 0.28 \\ 0.10 & 0.22 & 0.67 & 0.01 \\ 0.38 & 0.32 & 0.25 & 0.05 \end{bmatrix}, \tag{68a}$$

For this channel, Condition (61b) is satisfied for all non-zero inputs x_2 but at the same time Condition (61a) is violated for all non-zero inputs x_1. As described in Remark 7 the covert capacity-key tradeoff region $\mathcal{R}^\star(k_1, k_2)$ does not depend on the key-rate k_1 but only on k_2. Figure 6 illustrates the capacity key-rate tradeoff regions for this channel at different key rates.

V. Summary and Conclusion

We determined the covert capacity-key tradeoff for discrete and memoryless single-user and multi-access channels, i.e., the largest covert rates that are achievable for given key rates. Previous results had only determined the key rates required at the largest possible covert rates, while our results allow to conclude how much key-rate is required at any desired covert rate.

Our results provide new insights into the relationships between the required key rates and the achievable covert rates. For single-user channels with binary inputs, the relationship between key-rate and largest achievable covert rate is linear for low key rates before it saturates at the covert capacity. For non-binary input channels the relationship is linear for small key rates only if the adversary is uniformly (over all nonzero inputs) stronger (in a Kullback-Leibler-divergence sense) than the legitimate Rx. In this case, binary signaling between the zero and a single non-zero input is optimal in the regime of small key rates. For larger key rates or if the adversary is not uniformly stronger than the legitimate Rx, the largest achievable rate grows sublinearly in the key-rate before it saturates at the covert capacity. Binary signaling is not necessarily optimal in these cases.

In this work, we further characterized the covert capacity-key tradeoff region for discrete memoryless multiple-access channels (DMMAC). We showed that similarly to the DMC, the behavior of these regions depends on whether the warden observes a stronger channel than the legitimate receiver for the different inputs. When the warden is consistently stronger over all inputs, binary signaling is optimal at low key rates, and no tradeoff exists between the users' covert rateseach scales linearly with its own key rate. At higher key rates, the tradeoff region grows sublinearly, introducing interdependencies between the users rates. When the warden is stronger for some inputs but weaker for others, the tradeoff region exhibits sublinear growth at all key rates and maintains an inherent rate tradeoff between users.

Our results for the MAC can easily be extended to scenarios with more than two transmitters and also to interference channels. For the latter extension, notice that in our achievability proof for the MAC, the single receiver decodes each of the two messages individually, see (83) and (85). The same decoding can thus also be implemented at the two distributed receivers in an interference channel.

Appendix A
Proof of Theorem 5

The theorem follows directly from the following proposition and by noting that the optimal values of the parameters β_1 and β_2 are either equal to 1 or make that equality holds in (71) or (72), respectively.

Proposition 7: Given a tuple of message and key rates (r_1, r_2, k_1, k_2), it is possible to find a sequence (in the blocklength n) of tuples $(m_1, m_2, p_1, p_2, g_1, g_2)$ and encoding/decoding functions $(\varphi_1^{(n)}, \varphi_2^{(n)}, g^{(n)})$

satisfying (51)–(53) and so that $D_n \leq \delta_n$ for sufficiently large blocklengths n, if and only if, there exists a pmf ψ over $\tilde{\mathcal{X}}$ and two numbers $\beta_1, \beta_2 \in [0,1]$ so that the following inequalities hold:

$$r_1 \leq \beta_1 \sqrt{2} \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \mathbb{D}_{\mathsf{Y}}(x_1, 0)}{\sqrt{\chi^2(\psi)}}, \tag{69}$$

$$r_2 \leq \beta_2 \sqrt{2} \frac{\sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(0, x_2) \mathbb{D}_{\mathsf{Y}}(0, x_2)}{\sqrt{\chi^2(\psi)}}, \tag{70}$$

$$k_1 \geq \beta_1 \sqrt{2} \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \left(\mathbb{D}_{\mathsf{Z}}(x_1, 0) - \mathbb{D}_{\mathsf{Y}}(x_1, 0) \right)}{\sqrt{\chi^2(\psi)}} \tag{71}$$

$$k_2 \geq \beta_2 \sqrt{2} \frac{\sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(0, x_2) \left(\mathbb{D}_{\mathsf{Z}}(0, x_2) - \mathbb{D}_{\mathsf{Y}}(0, x_2) \right)}{\sqrt{\chi^2(\psi)}}. \tag{72}$$

Proof: We prove achievability followed by the converse.

Achievability:

Assume that $\beta_1 \geq \beta_2$, otherwise switch the roles of the two users. Set $\phi_j = \beta_j^2$, for $j \in \{1, 2\}$. In our coding scheme, Tx 1 communicates during the first

$$n_1 \triangleq \lfloor \phi_1 n \rfloor \tag{73}$$

channel uses and deterministically sends the all-zero sequence during the last $n - n_1$ channel uses. Tx 2 communicates only during the first

$$n_2 \triangleq \lfloor \phi_2 n \rfloor \leq n_1 \tag{74}$$

channel uses; during the following $n_1 - n_2$ channel uses it sends completely random symbols according to a distribution that we shall specify shortly; and during the last $n - n_1$ channel uses it sends the all-zero sequence. In our scheme Tx 2 thus uses its local randomness only if $\beta_1 > \beta_2$; and Tx 1 uses its local randomness when $\beta_2 > \beta_1$. Figure 7 illustrates this scheme. We next explain the generation of the codewords $x_1^{n_1}(W_1, S_1)$ and $x_2^{n_2}(W_2, S_2)$ and of the random inputs at Tx 2.

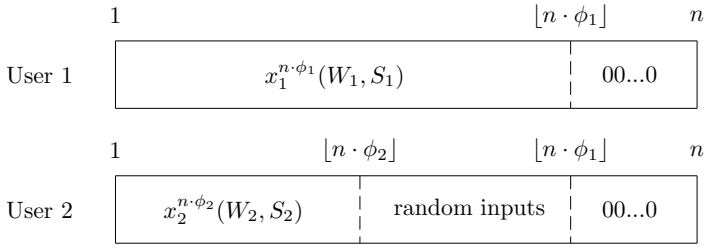

Fig. 7: Covert coding scheme for the MAC.

Let $\mu > 0$ be an arbitrary small number and define

$$\omega_n \triangleq \sqrt{\frac{2\delta_n}{(1+\mu) \cdot \phi_1 \cdot \chi^2(\psi)}}. \tag{75}$$

and

$$\alpha_n \triangleq \frac{\omega_n}{\sqrt{n}}. \tag{76}$$

Under our assumptions (11) we have

$$\lim_{n\to\infty} \omega_n = 0, \tag{77a}$$

$$\lim_{n\to\infty} \left(\omega_n\sqrt{n} - \log n\right) = \infty. \tag{77b}$$

We further pick a pmf ψ over the previously defined set $\tilde{\mathcal{X}} = ((\mathcal{X}_1\backslash\{0\}) \times \{0\}) \cup (\{0\} \times (\mathcal{X}_1\backslash\{0\}))$ and also define the pmfs (which depend on n)

$$P_{X_1}(x_1) = \psi(x_1, 0) \cdot \alpha_n, \quad \forall x_1 \in \mathcal{X}_1\backslash\{0\}, \tag{78}$$

$$P_{X_2}(x_2) = \psi(0, x_2) \cdot \alpha_n, \quad \forall x_2 \in \mathcal{X}_2\backslash\{0\}, \tag{79}$$

$$P_{X_j}(0) = 1 - \sum_{x_j\in\mathcal{X}_j\backslash\{0\}} P_{X_j}(x_j), \quad j \in \{1, 2\}, \tag{80}$$

$$P_{X_1X_2Y}(x_1, x_2, y) \triangleq P_{X_1}(x_1)P_{X_2}(x_2)\Gamma(y|x_1, x_2), \quad \forall (x_1, x_2, y) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}. \tag{81}$$

Codebook generation: For each $i = 1, \ldots, n_1$, independently draw the i-th entry of each codeword $x_1^{n_1}(w_1, s_1)$ according to P_{X_1}. For each $i = 1, \ldots, n_2$, independently draw the i-th entry of each codeword $x_2^{n_2}(w_2, s_2)$ according to P_{X_2}.

The codebooks are revealed to all parties.

Decoding at the Receiver: For any blocklength n' and positive constant η, define the set

$$\mathcal{A}_\eta^{n'} \triangleq \left\{(x_1^{n'}, x_2^{n'}, y^{n'}) \colon \log\left(\frac{\Gamma^{\otimes n'}(y^{n'}|x_1^{n'}, x_2^{n'})}{\Gamma^{\otimes n'}(y^{n'}|0^{n'}, 0^{n'})}\right) \geq \eta\right\}. \tag{82}$$

The legitimate Rx observes $Y^n = y^n$ and knows the secret-keys (S_1, S_2). To decode message W_1, it looks for a unique index w_1 satisfying

$$(x_1^{n_1}(w_1, S_1), 0^{n_1}, y^{n_1}) \in \mathcal{A}_{\eta_1}^{n_1}, \tag{83}$$

where for any $n' < n$ we let $y^{n'}$ denote the first n' symbols of y^n and

$$\eta_1 \triangleq (1-\mu)\phi_1\sqrt{n}\omega_n \cdot \sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1, 0)\mathbb{D}_{\mathsf{Y}}(x_1, 0). \tag{84}$$

If such a unique index w_1 exists, the legitimate Rx sets $\widehat{W}_1 = w_1$. Otherwise, it declares an error and stops.

Similarly, to decode message W_2, the legitimate Rx looks for a unique index w_2 satisfying

$$(0^{n_2}, x_2^{n_2}(w_2, S_2), y^{n_2}) \in \mathcal{A}_{\eta_2}^{n_2}, \tag{85}$$

for

$$\eta_2 \triangleq (1-\mu)\phi_2\sqrt{n}\omega_n \cdot \sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0, x_2)\mathbb{D}_{\mathsf{Y}}(0, x_2). \tag{86}$$

If such a unique index w_2 exists, it sets $\widehat{W}_2 = w_2$. Otherwise, it declares an error.

Performance Analysis: The first analysis steps are standard for covert communication and can be obtained, for example, by converting the steps in [23, Section IV] to our setup where message W_1 is decoded based on the key and the first n_1 outputs of Y^n and message W_2 based on the keys and the first n_2 outputs of Y^n. By these analysis steps, we obtain that for sufficiently large blocklengths n, we can choose

$$m_1 = (1-\mu)\phi_1 \cdot \omega_n\sqrt{n} \cdot \sum_{x_1\in\mathcal{X}_1\backslash\{0\}} \psi(x_1, 0)\mathbb{D}_{\mathsf{Y}}(x_1, 0), \tag{87a}$$

$$m_2 = (1-\mu)\phi_2 \cdot \omega_n\sqrt{n} \cdot \sum_{x_2\in\mathcal{X}_2\backslash\{0\}} \psi(0, x_2)\mathbb{D}_{\mathsf{Y}}(0, x_2), \tag{87b}$$

while guaranteeing

$$\Pr[\widehat{W}_j \neq W_j] \leq e^{-B\phi_j\cdot\omega_n\sqrt{n}}, \quad j \in \{1, 2\}, \tag{88}$$

for a positive real number B.

The resolvability analysis requires more care. Recall the definition of the warden's output distribution $\widehat{Q}(z^n)$ in (49) and that covertness is measured as:

$$D_n = \mathbb{D}(\widehat{Q}^n \| \Gamma^{\otimes n}(\cdot|0^n, 0^n)). \tag{89}$$

Define the pmfs

$$Q_{Z|X_1}(z|x_1) \triangleq \sum_{x_2 \in \mathcal{X}_2} P_{X_2}(x_2) Q(z|x_1, x_2), \quad z \in \mathcal{Z}, x_1 \in \mathcal{X}_1, \tag{90}$$

$$Q_{Z|X_1}(z|x_2) \triangleq \sum_{x_1 \in \mathcal{X}_1} P_{X_1}(x_1) Q(z|x_1, x_2), \quad z \in \mathcal{Z}, x_2 \in \mathcal{X}_2, \tag{91}$$

and

$$\tilde{Q}(z) \triangleq \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} P_{X_1}(x_1) P_{X_2}(x_2) Q(z|x_1, x_2), \quad z \in \mathcal{Z}. \tag{92}$$

Notice that D_n only depends on the first n_1 channel uses, as the terms corresponding to the last $n - n_1$ channel uses are zero. By Pinsker's inequality, we have the following bound:

$$\begin{aligned} &\left| D_n - \mathbb{D}\left(\tilde{Q}^{\otimes n_1} \,\|\, Q^{\otimes n_1}(\cdot|0^{n_1}, 0^{n_1})\right) \right| \\ &\leq \mathbb{D}\left(\widehat{Q}^{n_1} \,\|\, \tilde{Q}^{\otimes n_1}\right) + n_1 \log\left(\frac{1}{\min_{z \in \mathcal{Z}} Q(z|0,0)}\right) \sqrt{\frac{1}{2}\mathbb{D}\left(\widehat{Q}^{n_1} \,\|\, \tilde{Q}^{\otimes n_1}\right)}, \end{aligned} \tag{93}$$

where $\widehat{Q}^{n_1}$ denotes the pmf of the warden's first n_1 outputs.

Since $\alpha_n \to 0$ as $n \to \infty$, by [17, Eq. (295)], we can write

$$\mathbb{D}\left(\tilde{Q}^{\otimes n_1} \,\|\, Q^{\otimes n_1}(\cdot|0^{n_1}, 0^{n_1})\right) = (1 + o(1)) \cdot n_1 \cdot \frac{\alpha_n^2}{2} \cdot \chi^2(\psi) \tag{94}$$

$$= (1 + o(1)) \cdot \frac{n_1}{n} \cdot \frac{\omega_n^2}{2} \cdot \chi^2(\psi) \tag{95}$$

$$= \frac{(1 + o(1))}{1 + \mu} \delta_n, \tag{96}$$

where in the last equality we used that the ratio $\frac{n_1}{n} \to \phi_1$ and the definition of ω_n in (75). By Lemma 2 ahead, the divergence term $\mathbb{D}\big(\widehat{Q}^{n_1} \,\|\, \tilde{Q}^{\otimes n_1}\big)$ vanishes exponentially in $\sqrt{n}\omega_n$. Combining this result with (93) and (96), we can conclude that our coding scheme satisfies the divergence constraint $D_n \leq \delta_n$ for all sufficiently large blocklengths n.

Using the expressions for the message bits in (87) and the definition of ω_n in (75), and letting $n \to \infty$ and $\mu, \xi_1, \xi_2, \xi_3 \downarrow 0$ (since $n_1 \approx n\phi_1$ and $n_2 \approx n\phi_2$), we conclude achievability of the message-key rate tuples (r_1, r_2, k_1, k_2)

$$r_1 = \sqrt{\phi_1} 2 \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \mathbb{D}_{\mathsf{Y}}(x_1, 0)}{\sqrt{\chi^2(\psi)}}, \tag{97}$$

$$r_2 = \frac{\phi_1 \sqrt{2}}{\sqrt{\phi_2}} \frac{\sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(0, x_2) \mathbb{D}_{\mathsf{Y}}(0, x_2)}{\sqrt{\chi^2(\psi)}}, \tag{98}$$

$$k_1 \geq \sqrt{\phi_1} 2 \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \left(\mathbb{D}_{\mathsf{Z}}(x_1, 0) - \mathbb{D}_{\mathsf{Y}}(x_1, 0)\right)}{\sqrt{\chi^2(\psi)}}, \tag{99}$$

$$k_2 \geq \frac{\phi_1\sqrt{2}}{\sqrt{\phi_2}} \frac{\sum\limits_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(0,x_2)\left(\mathbb{D}_{\mathsf{Z}}(0,x_2) - \mathbb{D}_{\mathsf{Y}}(0,x_2)\right)}{\sqrt{\chi^2(\psi)}}. \tag{100}$$

Setting $\beta_1 = \sqrt{\phi_1}$ and $\beta_2 = \frac{\phi_2}{\sqrt{\phi_1}}$ and remarking that we can relax the equalities on the message rates into $\leq$-inequalities (we can always decide to send dummy information bits) proves the desired achievability result.

Lemma 2: Let

$$\bar{\theta}_1 \triangleq \alpha_n n_1 \cdot \sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1,0)\mathbb{D}_{\mathsf{Z}}(x_1,0), \tag{101a}$$

$$\bar{\theta}_2 \triangleq \alpha_n n_2 \cdot \sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(0,x_2)\mathbb{D}_{\mathsf{Z}}(0,x_2), \tag{101b}$$

and $\theta_i = (1+\xi_1)\bar{\theta}_i$, for $i = 1,2$ and arbitrary small but positive constants ξ_1, ξ_2. If the message and key bits satisfy

$$\limsup_{n\to\infty} (m_1 + p_1 - \theta_1) = -\infty, \tag{102a}$$

$$\limsup_{n\to\infty} (m_2 + p_2 - \theta_2) = -\infty, \tag{102b}$$

then $\mathbb{D}\big(\widehat{Q}^{n_1} \,\|\, \tilde{Q}^{\otimes n_1}\big)$ tends to 0 as $n \to \infty$ at a speed that is exponential in $\omega_n\sqrt{n}$.

Proof: Define for a given codebook $\mathcal{C}$:

$$\widehat{Q}^{n_1}(z^{n_1}) \triangleq \frac{1}{2^{m_1+m_2+p_1+p_2}} \sum_{(w_1,s_1)} \sum_{(w_2,s_2)} \widehat{Q}^{n_1}_{w_1,w_2,s_1,s_2}(z^{n_1}), \tag{103}$$

where for any valid (w_1,w_2,s_1,s_2), we have:

$$\widehat{Q}^{n_1}_{w_1,w_2,s_1,s_2}(z^{n_1}) \triangleq Q^{\otimes n_2}(z^{n_2}|x_1^{n_2}(w_1,s_1), x_2^{n_2}(w_2,s_2)) \cdot Q^{(n_2\to n_1)}_{Z|X_1}(z_{n_2+1}^{n_1}|x_{1,n_2+1}^{n_1}(w_1,s_1)), \tag{104}$$

for $z_{n_2+1}^{n_1} \triangleq (z_{n_2+1},\ldots,z_{n_1})$ and for $Q^{(n_2\to n_1)}_{Z|X_1}$ defined as:

$$Q^{(n_2\to n_1)}_{Z|X_1}\left(z_{n_2+1}^{n_1}|x_{1,n_2+1}^{n_1}(w_1,s_1)\right) \triangleq \prod_{i=n_2+1}^{n_1} Q_{Z|X_1}(z_i|x_{1,i}). \tag{105}$$

We can now notice:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[\mathbb{D}\left(\widehat{Q}^{n_1}\middle\|\tilde{Q}^{\otimes n_1}\right)\right] \\
&\leq \mathbb{E}\left[\log\left(\sum_{(w_1,w_2,s_1,s_2)} \mathbb{E}_{\substack{\{X_1^{n_1}(w_1,s_1)\}\setminus X_1^{n_1}(1,1),\\ \{X_2^{n_2}(w_2,s_2)\}\setminus X_2^{n_2}(1,1)}} \left[\frac{\widehat{Q}^{n_1}_{w_1,w_2,s_1,s_2}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})}\right]\right)\right] &\text{(106a)}\\
&= \mathbb{E}\Bigg[\log\Bigg(\sum_{\substack{(w_1,s_1)\neq(1,1)\\(w_2,s_2)\neq(1,1)}} \mathbb{E}_{\substack{X_1^{n_1}(w_1,s_1)\\X_2^{n_2}(w_2,s_2)}} \left[\frac{\widehat{Q}^{n_1}_{w_1,w_2,s_1,s_2}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})}\right] + \frac{\widehat{Q}^{n_1}_{1,1,1,1}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} \\
&\qquad + \sum_{(w_2,s_2)\neq(1,1)} \mathbb{E}_{X_2^{n_2}(w_2,s_2)} \left[\frac{\widehat{Q}^{n_1}_{1,w_2,1,s_2}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})}\right] \\
&\qquad + \sum_{(w_1,s_1)\neq(1,1)} \mathbb{E}_{X_1^{n_1}(w_1,s_1)} \left[\frac{\widehat{Q}^{n_1}_{w_1,1,s_1,1}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})}\right]\Bigg)\Bigg] &\text{(106b)}
\end{aligned}$$

$$\stackrel{(a)}{=} \mathbb{E}\Bigg[\log\Bigg(\frac{(2^{m_1+p_1}-1)(2^{m_2+p_2}-1)}{2^{m_1+m_2+p_1+p_2}} + \frac{\widehat{Q}^{n_1}_{1,1,1,1}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{(2^{m_2+p_2}-1)Q^{\otimes n_1}_{Z|X_1}(Z^{n_1}|X_1^{n_1}(1,1))}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{(2^{m_1+p_1}-1)Q^{\otimes n_2}_{Z|X_2}(Z^{n_2}|X_2^{n_2}(1,1))}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_2}(Z^{n_2})}\Bigg)\Bigg] \tag{106c}$$

$$\leq \mathbb{E}\Bigg[\log\Bigg(1 + \frac{\widehat{Q}^{n_1}_{1,1,1,1}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{Q^{n_1}_{Z|X_1}(Z^{n_1}|X_1^{n_1}(1,1))}{2^{m_1+p_1}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{Q^{n_2}_{Z|X_2}(Z^{n_2}|X_2^{n_2}(1,1))}{2^{m_2+p_2}\cdot\tilde{Q}^{\otimes n_2}(Z^{n_2})}\Bigg)\Bigg]. \tag{106d}$$

Define for any pair $\boldsymbol{\theta} = (\theta_1, \theta_2)$ the set:

$$\begin{aligned}\mathcal{B}^{n_1}_{\boldsymbol{\theta}} \triangleq \Bigg\{&(x_1^{n_1}, x_2^{n_2}, z^{n_1}) \in \mathcal{X}_1^{n_1}\times\mathcal{X}_2^{n_2}\times\mathcal{Z}^{n_1}:\\ &\log\left(\frac{Q^{\otimes n_2}(z^{n_2}|x_1^{n_2}, x_2^{n_2})\cdot Q^{n_2\to n_1}_{Z|X_1}(z^{n_1}_{n_2+1}|x^{n_1}_{1,n_2+1})}{Q^{\otimes n_1}(z^{n_1}|0^{n_1}, 0^{n_1})}\right) \leq \theta_1+\theta_2,\\ &\log\left(\frac{Q^{n_1}_{Z|X_1}(z^{n_1}|x_1^{n_1})}{Q^{\otimes n_1}(z^{n_1}|0^{n_1}, 0^{n_1})}\right) \leq \theta_1,\\ &\log\left(\frac{Q^{n_2}_{Z|X_2}(z^{n_2}|x_2^{n_2})}{Q^{\otimes n_2}(z^{n_2}|0^{n_2}, 0^{n_2})}\right) \leq \theta_2\Bigg\}.\end{aligned} \tag{107}$$

and notice that $x_2^{n_2}$ is of length n_2 while $x_1^{n_1}$ and $z_1^{n_1}$ are of length n_1. Further define the event

$$\mathcal{B} \triangleq \{(X_1^{n_1}(W_1,S_1), X_2^{n_2}(W_2,S_2), Z^{n_1}) \in \mathcal{B}^{n_1}_{\boldsymbol{\theta}}\} \tag{108}$$

and denote its complement by $\mathcal{B}^c$.

We continue to bound

$$\mathbb{E}\Bigg[\log\Bigg(1 + \frac{\widehat{Q}^{n_1}_{1,1,1,1}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{\Gamma^{n_1}_{Z|X_1}(Z^{n_1}|X_1^{n_1}(1,1))}{2^{m_1+p_1}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{\Gamma^{n_2}_{Z|X_2}(Z^{n_2}|X_2^{n_2}(1,1))}{2^{m_2+p_2}\cdot\tilde{Q}^{\otimes n_2}(Z^{n_2})}\Bigg)\Bigg|\mathcal{B}\Bigg]\Pr[\mathcal{B}]$$

$$\stackrel{(a)}{\leq} \mathbb{E}\Bigg[\log\Bigg(1 + \frac{2^{\theta_1+\theta_2}Q^{\otimes n_1}(Z^{n_1}|0^{n_1}, 0^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1}))} + \frac{2^{\theta_1}Q^{\otimes n_1}(Z^{n_1}|0^{n_1}, 0^{n_1})}{2^{m_1+p_1}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{2^{\theta_2}Q^{\otimes n_2}(Z^{n_2}|0^{n_2}, 0^{n_2})}{2^{m_2+p_2}\cdot\tilde{Q}^{\otimes n_2}(Z^{n_2})}\Bigg)\Bigg|\mathcal{B}\Bigg]\Pr[\mathcal{B}] \tag{109a}$$

$$\leq \frac{2^{\theta_1+\theta_2}}{2^{m_1+m_2+p_1+p_2}}\mathbb{E}\left[\frac{Q^{\otimes n_1}(Z^{n_1}|0^{n_1}, 0^{n_1})}{\tilde{Q}^{\otimes n_1}(Z^{n_1})}\right] + \frac{2^{\theta_1}}{2^{m_1+p_1}}\mathbb{E}\left[\frac{Q^{\otimes n_1}(Z^{n_1}|0^{n_1}, 0^{n_1})}{\tilde{Q}^{\otimes n_1}(Z^{n_1})}\right] + \frac{2^{\theta_2}}{2^{m_2+p_2}}\mathbb{E}\left[\frac{Q^{\otimes n_2}(Z^{n_2}|0^{n_2}, 0^{n_2})}{\tilde{Q}^{\otimes n_2}(Z^{n_2})}\right] \tag{109b}$$

$$\stackrel{(b)}{=} \frac{2^{\theta_1+\theta_2}}{2^{m_1+m_2+p_1+p_2}} + \frac{2^{\theta_1}}{2^{m_1+p_1}} + \frac{2^{\theta_2}}{2^{m_2+p_2}}, \tag{109c}$$

where (a) holds by the definition of the set $\mathcal{B}^n_{\boldsymbol{\theta}}$ in (107); and (b) holds because $Z^{n'} \sim \tilde{Q}^{\otimes n'}$ for any n' and because pmfs sum to 1.

Moreover, since by definition (92), we can lower bound $\tilde{Q}(z)$ by $P_{X_1}(0)P_{X_2}(0)\min_{z\in\mathcal{Z}} Q(z|0,0)$, we obtain:

$$\mathbb{E}\Bigg[\log\Bigg(1 + \frac{\widehat{Q}^{n_1}_{1,1,1,1}(Z^{n_1})}{2^{m_1+m_2+p_1+p_2}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{Q^{n_1}_{Z|X_1}(Z^{n_1}|X_1^{n_1}(1,1))}{2^{m_1+p_1}\cdot\tilde{Q}^{\otimes n_1}(Z^{n_1})} + \frac{Q^{n_2}_{Z|X_2}(Z^{n_2}|X_2^{n_2}(1,1))}{2^{m_2+p_2}\cdot\tilde{Q}^{\otimes n_2}(Z^{n_2})}\Bigg)\Bigg|\mathcal{B}^c\Bigg]$$

$$\leq n_1 \log\left(\frac{4}{\left(1-\sum_{x_1\in\mathcal{X}_1\setminus\{0\}}\psi(x_1,0)\alpha_n\right)\left(1-\sum_{x2\in\mathcal{X}_2\setminus\{0\}}\psi(0,x_2)\alpha_n\right)\cdot\min_{z\in\mathcal{Z}}Q(z|0,0)}\right). \quad (110)$$

To bound the probability of event $\mathcal{B}^c$, we notice that the expectations of the log-terms in (107) can be upper bounded, respectively, by $\bar{\theta}_1+\bar{\theta}_2, \bar{\theta}_1$, and $\bar{\theta}_2$ plus some functions $\mathcal{O}(\alpha_n^2)$ that vanish in the order of α_n^2. Since also the variances of above log-expressions vanish in the order of α_n, we can apply Bernstein's Inequality and the union bound to conclude that by our choice $\theta_i=(1+\xi_i)\bar{\theta}_i$, for $i=1,2,$, there exist positive constants $B_1,B_2,B_3>0$ such that

$$\Pr[\mathcal{B}^c]\leq e^{-B_1\omega_n\sqrt{n}}+e^{-B_2\omega_n\sqrt{n}}+e^{-B_3\omega_n\sqrt{n}} \quad (111)$$

Combining finally (106d) with (109c), (110), and (111) through the total law of expectation, we conclude that under Conditions (102), the divergence $\mathbb{D}\left(\widehat{Q}^{n_1}\,\|\,\tilde{Q}^{\otimes n_1}\right)$ tends to 0 exponentially fast in $\sqrt{n}\omega_n$. ■

Converse:

The particularity and the distinction of our converse proof from the previously established one [17, Section V.C] is that users are equipped with local randomness. We highlight the main differences.

Consider a vanishing sequence $\{\delta_n\}$ and a sequence of length-n codes with vanishing probability of error and $D_n\leq\delta_n$ for all sufficiently large blocklengths. Consider now a fixed blocklength n, and let X_1^n, X_2^n be the random inputs generated under the chosen codes and Y^n as well as Z^n the corresponding outputs at the legitimate Rx and the warden.

Define

$$\alpha_{n,j,i}(x)\triangleq\Pr[X_{j,i}=x],\quad i\in\{1,\ldots,n\},\ \forall j\in\{1,2\}, \quad (112)$$

and the derived nonnegative quantities

$$\alpha_n\triangleq\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{x_1\in\mathcal{X}_1\setminus\{0\}}\alpha_{n,1,i}(x_1)+\sum_{x_2\in\mathcal{X}_2\setminus\{0\}}\alpha_{n,2,i}(x_2)\right), \quad (113)$$

and

$$\psi_n(x_1,0)=\frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,1,i}(x_1)}{\alpha_n}, \quad (114)$$

$$\psi_n(0,x_2)=\frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,2,i}(x_2)}{\alpha_n}. \quad (115)$$

Notice that when the probability of decoding errors tends to 0 and the message sizes m_1 and m_2 tend to ∞, the sequence α_n cannot decrease to 0 as $1/n$ because otherwise in the limit most codewords consist only of the 0 symbol and cannot be distinguished. We can thus assume in the sequel that $\lim_{n\to\infty}n\alpha_n=\infty$.

Lower bound on δ_n: We start by writing:

$$D_n=\sum_{z^n}\widehat{Q}^n(z^n)\log\left(\frac{\widehat{Q}^n(z^n)}{Q^{\otimes n}(z^n|0^n,0^n)}\right) \quad (116)$$

$$\overset{(a)}{\geq}\sum_{i=1}^{n}\sum_{z_i}\widehat{Q}^{(i)}(z_i)\log\left(\frac{\widehat{Q}^{(i)}(z_i)}{Q(z_i|0,0)}\right) \quad (117)$$

$$=\sum_{i=1}^{n}\mathbb{D}\left(\widehat{Q}^{(i)}\|Q(\cdot|0,0)\right) \quad (118)$$

$$\overset{(b)}{=}\sum_{i=1}^{n}\mathbb{D}\left(\sum_{\substack{x_1\in\mathcal{X}_1\\x_2\in\mathcal{X}_2}}\alpha_{n,1,i}(x_1)\alpha_{n,2,i}(x_2)Q(\cdot|x_1,x_2)\,\Big\|\,Q(\cdot|0,0)\right) \quad (119)$$

$$\overset{(c)}{\geq} n\mathbb{D}(\bar{Q}\|Q(\cdot|0,0)) \quad (120)$$

$$\overset{(d)}{=} n \sum_{z\in\mathcal{Z}} Q(z|0,0)\left(1+\frac{\alpha_n\zeta_n(z)}{Q(z|0,0)}\right)\log\left(1+\frac{\alpha_n\zeta_n(z)}{Q(z|0,0)}\right) \tag{121}$$

$$\overset{(e)}{\geq} n\frac{\alpha_n^2}{2}\sum_{z\in\mathcal{Z}}\frac{\zeta_n^2(z)}{Q(z|0,0)} \tag{122}$$

where above sequence of (in)equalities and approximations are justified as follows:

- (a) holds by the memoryless nature of the channel, by the convexity of the $t\mapsto t\log t$ function, and by defining $\widehat{Q}_{\mathcal{C}}^{(i)}(z_i)$ as the probability of the event $Z_i=z_i$;
- (b) by writing out the expectations over the independent random variables $X_{1,i}$ and $X_{2,i}$;
- (c) holds by the convexity of the relative entropy and

$$\bar{Q}(z)\triangleq\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{x_1\in\mathcal{X}_1\\x_2\in\mathcal{X}_2}}\alpha_{n,1,i}(x_1)\alpha_{n,2,i}(x_2)Q(\cdot|x_1,x_2); \tag{123}$$

- (d) holds by expanding the KL divergence term and upon noticing that we can rewrite $\bar{Q}$ as

$$\bar{Q}(z)=Q(z|0,0)\left(1+\frac{\alpha_n\left[\bar{Q}(z)-Q(z|0,0)\right]}{\alpha_n Q(z|0,0)}\right), \tag{124}$$

  and defining

$$\zeta_n(z)\triangleq\frac{\bar{Q}(z)-Q(z|0,0)}{\alpha_n}. \tag{125}$$

- (e) follows from the inequalities $\log(1+x)>x-\frac{x^2}{2}$ for $x\geq 0$ and $\log(1+x)>x-\frac{x^2}{2}+\frac{2x^3}{3}$ for $x\in[\frac{-1}{2},0]$ and by loosening the bounds because for large n, we have $\alpha_n\ll 1$ and thus $\alpha_n^4\ll\alpha_n^3\ll\alpha_n^2$.

Since $\lim_{n\to\infty}\delta_n=0$, by (119) we can conclude that

$$\lim_{n\to\infty}\alpha_{n,j,i}(x_j)=0,\quad \forall i\in\{1,\dots,n\},\ j\in\{1,2\},\ \forall x_j\in\mathcal{X}_j\setminus\{0\}, \tag{126}$$

and hence

$$\lim_{n\to\infty}\alpha_{n,j,i}(0)=1,\quad \forall i\in\{1,\dots,n\},\ j\in\{1,2\}. \tag{127}$$

We now derive the asymptotics of ζ_n which are going to serve later when we take limits. Consider a subsequence of blocklengths $\{n_\ell\}_{\ell=1}^{\infty}$ for which all expressions

$$\psi_n(x_1,0)\triangleq\frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,1,i}(x_1)}{\alpha_n} \tag{128a}$$

$$\psi_n(0,x_2)\triangleq\frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,2,i}(x_2)}{\alpha_n} \tag{128b}$$

converge, and denote the respective convergence points by

$$\psi(x_1,0)\triangleq\lim_{\ell\to\infty}\frac{\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\alpha_{n_\ell,1,i}(x_1)}{\alpha_{n_\ell}} \tag{129a}$$

$$\psi(0,x_2)\triangleq\lim_{\ell\to\infty}\frac{\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\alpha_{n_\ell,2,i}(x_2)}{\alpha_{n_\ell}}. \tag{129b}$$

Notice that ψ forms a pmf over the alphabet $\tilde{\mathcal{X}}$. It can then be shown that:

$$\zeta(z):=\lim_{\ell\to\infty}\zeta_{n_\ell}(z) \tag{130}$$

$$=\sum_{(x_1,x_2)\in\tilde{\mathcal{X}}}\psi(x_1,x_2)\big(Q(z|x_1,x_2)-Q(z|0,0)\big). \tag{131}$$

Since $D_n \leq \delta_n$ for sufficiently large blocklengths n, we can conclude that for sufficiently large ℓ:

$$\delta_{n_\ell} \geq (1+o(1)) \cdot n_\ell \frac{\alpha_{n_\ell}^2}{2} \sum_{z \in \mathcal{Z}} \frac{\zeta^2(z)}{Q(z|0,0)}. \tag{132}$$

Upper bound on m_1 and m_2: For a given blocklength n, let

$$P_e \triangleq \Pr\left[\hat{W}_1 \neq W_1 \text{ or } \hat{W}_2 \neq W_2\right] \tag{133}$$

Since the message W_1 is uniform over $\{1, \ldots, \mathsf{M}_1\}$ and independent of the local randomness C_1, C_2, we have

$$m_1 \leq \frac{1}{1-P_e}(\mathbb{I}(W_1; Y^n | W_2, S_1, S_2, C_1, C_2, X_2^n) + \mathbb{H}_b(P_e)) \tag{134}$$

$$\leq \frac{1}{1-P_e} \cdot \sum_{i=1}^{n} \left(\mathbb{H}(Y_i|X_{2i}) - \mathbb{H}(Y_i|X_{1i}, X_{2i}) + \mathbb{H}_b(P_e)\right) \tag{135}$$

$$= \frac{1}{1-P_e} \left(n \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}(X_{1i}; Y_i | X_{2i}) + \mathbb{H}_b(P_e) \right). \tag{136}$$

Define T as the time sharing random variable which is uniform over $\{1, \ldots, n\}$ and independent of all other random variables. Then:

$$m_1 \leq \frac{1}{1-P_e} \left(n \mathbb{I}(\tilde{X}_1; \tilde{Y} | \tilde{X}_2, T) + 1 \right) \tag{137}$$

$$\leq \frac{n(1+o(1))}{1-P_e} \left(\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} P_{X_{1,T}}(x_1) \cdot \mathbb{D}_{\mathsf{Y}}(x_1, 0) + \frac{1}{n} \right), \tag{138}$$

$$= \frac{n\alpha_n(1+o(1))}{1-P_e} \cdot \sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi_n(x_1, 0) \mathbb{D}_{\mathsf{Y}}(x_1, 0), \tag{139}$$

where the second inequality is obtained by applying [17, Lemma 1] for each realization of T and by noticing the joint pmf

$$P_{X_{1,T} X_{2,T} Y_T T}(x_1, x_2, y, t) = P_T(t) P_{X_{1,T}|T}(x_1|t) P_{X_{2,T}|T}(x_2|t) \Gamma(y|x_1, x_2) \tag{140}$$

for P_T uniform over $\{1, \ldots, n\}$; and the last equality holds by using the fact that $\alpha_n n$ cannot tend to 0 because otherwise no information is transmitted. Similar steps can be used to bound m_2.

Since $P_e \to 0$ as $n \to \infty$ and ψ_n converges on the sequence of blocklengths $\{n_\ell\}$, we obtain for each n_ℓ:

$$m_1 \leq n_\ell \alpha_{n_\ell}(1+o(1)) \sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \mathbb{D}_{\mathsf{Y}}(x_1, 0). \tag{141}$$

$$m_2 \leq n_\ell \alpha_{n_\ell}(1+o(1)) \sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(0, x_2) \mathbb{D}_{\mathsf{Y}}(0, x_2). \tag{142}$$

Upper bound on r_1 and r_2: Combining (141) with (132), we obtain:

$$\lim_{\ell \to \infty} \frac{m_1}{\sqrt{n_\ell \delta_{n_\ell}}} \leq \sqrt{2} \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \cdot \mathbb{D}_{\mathsf{Y}}(x_1, 0)}{\sqrt{\sum_{z \in \mathcal{Z}} \frac{\zeta^2(z)}{Q(z|0,0)}}} \tag{143}$$

$$= \sqrt{2} \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \cdot \mathbb{D}_{\mathsf{Y}}(x_1, 0)}{\sqrt{\chi^2(\psi)}}, \tag{144}$$

and

$$\lim_{\ell\to\infty}\frac{m_2}{\sqrt{n_\ell\delta_{n_\ell}}}\leq\sqrt{2}\frac{\sum_{x_2\in\mathcal{X}_2\setminus\{0\}}\psi(0,x_2)\cdot\mathbb{D}_{\mathsf{Y}}(0,x_2)}{\sqrt{\chi^2(\psi)}}. \tag{145}$$

Upper bound on δ_n: Let β_1 and β_2 be the two numbers in $[0,1]$ that satisfy

$$\lim_{\ell\to\infty}\frac{m_1}{\sqrt{n_\ell\delta_{n_\ell}}}=\beta_1\sqrt{2}\frac{\sum_{x_1\in\mathcal{X}_1\setminus\{0\}}\psi(x_1,0)\cdot\mathbb{D}_{\mathsf{Y}}(x_1,0)}{\sqrt{\chi^2(\psi)}} \tag{146a}$$

$$\lim_{\ell\to\infty}\frac{m_2}{\sqrt{n_\ell\delta_{n_\ell}}}=\beta_2\sqrt{2}\frac{\sum_{x_2\in\mathcal{X}_2\setminus\{0\}}\psi(0,x_2)\cdot\mathbb{D}_{\mathsf{Y}}(0,x_2)}{\sqrt{\chi^2(\psi)}}. \tag{146b}$$

Assume for the moment that β_1,β_2 are strictly larger than 0, and thus we can divide by them. Then, (146) combined with (141) and (142) implies that for all blocklengths n_ℓ:

$$\sqrt{n_\ell\delta_{n_\ell}}\leq\frac{n_\ell\alpha_{n_\ell}}{\beta_j\sqrt{2}(1-P_e)}\sqrt{\chi^2(\psi)}(1+o(1)),\quad j\in\{1,2\}. \tag{147}$$

Lower bound on m_j+p_j, for $j\in\{1,2\}$: We start with the lower bound

$$\begin{aligned}m_1+p_1&=\mathbb{H}(W_1,S_1|C_1,C_2) &(148)\\ &\geq\mathbb{I}(W_1,S_1;Z^n|C_1,C_2) &(149)\\ &\overset{(a)}{\geq}\mathbb{I}(X_1^n;Z^n|C_1,C_2) &(150)\\ &\overset{(b)}{\geq}\mathbb{I}(X_1^n,X_2^n;Z^n|C_1,C_2)-\mathbb{I}(X_2^n;Z^n|X_1^n) &(151)\end{aligned}$$

where above sequence of (in)equalities are justified as follows:

- (a) holds because $X_1^n=x_1^n(W_1,S_1,C_1)$ is a function of W_1, S_1, and C_1.
- (b) holds because of the Markov chain $(C_1,C_2)\to(X_1^n,X_2^n)\to Z^n$ and because conditioning reduces entropy.

Similarly we get,

$$m_2+p_2\geq\mathbb{I}(X_1^n,X_2^n;Z^n|C_1,C_2)-\mathbb{I}(X_1^n;Z^n|X_1^n). \tag{152}$$

We next focus on the first mutual-information term that is common to the RHS of (151) and (152). To this end, we define for each pair $(c_1,c_2)\in\mathcal{G}_1\times\mathcal{G}_2$ the warden's average output distribution:

$$\widehat{Q}^n_{(c_1,c_2)}(z^n)\triangleq\frac{\sum_{(w_1,s_1)}\sum_{(w_2,s_2)}Q^{\otimes n}(z^n|x_1^n(w_1,s_1,c_1),x_2^n(w_2,s_2,c_2))}{2^{m_1+m_2+p_1+p_2}} \tag{153}$$

and the divergence

$$D_{n,(c_1,c_2)}\triangleq\mathbb{D}\left(\widehat{Q}^n_{(c_1,c_2)}\,\middle\|\,Q^{\otimes n}(\cdot|0^n,0^n)\right). \tag{154}$$

With these definitions, we can write:

$$\begin{aligned}&\mathbb{I}(X_1^n,X_2^n;Z^n|C_1,C_2) &(155)\\ &\overset{(a)}{=}\mathbb{E}\left[\sum_{z^n}Q^{\otimes n}(z^n|X_1^n,X_2^n)\log\left(\frac{Q^{\otimes n}(z^n|X_1^n,X_2^n)}{Q^{\otimes n}(z^n|0^n,0^n)}\right)\right]-\mathbb{E}\left[D_{n,(C_1,C_2)}\right] &(156)\\ &\overset{(b)}{\geq}\sum_{i=1}^n\mathbb{E}\left[\sum_{z_i}Q(z_i|X_{1,i},X_{2,i})\log\left(\frac{Q(z_i|X_{1i},X_{2i})}{Q(z_i|0,0)}\right)\right]-D_n &(157)\end{aligned}$$

$$\stackrel{(c)}{=} \sum_{i=1}^{n} \left(\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \alpha_{n,1,i}(x_1) \mathbb{D}_{\mathsf{Z}}(x_1, 0) + \sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \alpha_{n,2,i}(x_2) \mathbb{D}_{\mathsf{Z}}(0, x_2) \right) (1 + o(1)) - D_n \tag{158}$$

$$\stackrel{(d)}{=} n\alpha_n \left(\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi_n(x_1, 0) \mathbb{D}_{\mathsf{Z}}(x_1, 0) + \sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi_n(0, x_2) \mathbb{D}_{\mathsf{Z}}(0, x_2) \right) (1 + o(1)) - D_n, \tag{159}$$

where above sequence of (in)equalities are justified as follows:

- (a) holds by the definition of $D_{n,(c_1,c_2)}$;
- (b) holds by convexity of the divergence;
- (c) by writing out the expectations over the independent random variables $X_{1,i}$ and $X_{2,i}$; by noting that for $X_{1,i} = X_{2,i} = 0$ the term in the expectation evaluates to 0; and by recalling that all $\alpha_{n,1,i}, \alpha_{n,2,i} \to 0$ as $n \to \infty$;
- (d) holds by the definition of ψ_n in (128).

For the second mutual-information term on the RHS of (151), we have:

$$\mathbb{I}(X_1^n; Z^n | X_2^n) \leq \sum_{i=1}^{n} \mathbb{H}(Z_i | X_{2,,i}) - \mathbb{H}(Z_i | X_{1,i}, X_{2,i}) \tag{160}$$

$$= n\mathbb{I}(X_{1,T}; Z_T | X_{2,T}, T) \tag{161}$$

$$\leq n\alpha_n(1 + o(1)) \cdot \sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi_n(x_1, 0) \mathbb{D}_{\mathsf{Z}}(x_1, 0), \tag{162}$$

where the last inequality holds by applying [17, Lemma 1] to outputs Z_T and for each realization of T. Similarly,

$$\mathbb{I}(X_2^n; Z^n | X_1^n) \leq n\alpha_n(1 + o(1)) \cdot \sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi_n(0, x_2) \mathbb{D}_{\mathsf{Z}}(0, x_2). \tag{163}$$

Thus, combining (159), (162) and (163) with (151)–(152) and $D_n \leq \delta_n$, for sufficiently large values of n we obtain:

$$m_1 + p_1 \geq n\alpha_n(1 + o(1)) \cdot \sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi_n(x_1, 0) \mathbb{D}_{\mathsf{Z}}(x_1, 0) - \delta_n \tag{164}$$

and

$$m_2 + p_2 \geq n\alpha_n(1 + o(1)) \cdot \sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi_n(0, x_2) \mathbb{D}_{\mathsf{Z}}(0, x_2) - \delta_n. \tag{165}$$

Lower bound on $r_j + k_j$, for $j \in \{1, 2\}$: Restricting to the subsequence of blocklengths $\{n_\ell\}$, by the definition of $\psi(x_1, 0)$ and $\psi(0, x_2)$, we obtain from (164), (165), and (147):

$$\lim_{\ell \to \infty} \frac{m_1 + p_1}{\sqrt{n_\ell \delta_{n_\ell}}} \geq \beta_1 \sqrt{2} \frac{\sum_{x_1 \in \mathcal{X}_1 \setminus \{0\}} \psi(x_1, 0) \mathbb{D}_{\mathsf{Z}}(x_1, 0)}{\sqrt{\chi^2(\psi)}} \tag{166}$$

and

$$\lim_{\ell \to \infty} \frac{m_2 + p_2}{\sqrt{n_\ell \delta_{n_\ell}}} \geq \beta_2 \sqrt{2} \frac{\sum_{x_2 \in \mathcal{X}_2 \setminus \{0\}} \psi(x_2, 0) \mathbb{D}_{\mathsf{Z}}(0, x_2)}{\sqrt{\chi^2(\psi)}}. \tag{167}$$

Here we used the fact that $\sqrt{\frac{\delta_n}{n}}$ vanishes for increasing blocklengths because $\delta_n \to 0$ as $n \to \infty$.

Concluding the Proof of the Proposition: The proof of the proposition is then concluded by combining (166) and (167) with (146), and by noting that this latter equality can be turned into an inequality because the Txs are always allowed to send dummy information bits. ■

REFERENCES

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

[2] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: fundamental limits of covert wireless communication," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 26–31, 2015.

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

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

[5] C. Bouette, L. Luzzi, and L. Wang, "Covert communication over additive-noise channels," *IEEE Transactions on Information Theory*, vol. 71, no. 3, pp. 2157–2169, 2025.

[6] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *2014 IEEE Information Theory Workshop (ITW 2014)*, 2014, pp. 30–34.

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

[8] H. ZivariFard, M. Bloch, and A. Nosratinia, "Keyless covert communication in the presence of non-causal channel state information," in *2019 IEEE Information Theory Workshop (ITW)*, 2019, pp. 1–5.

[9] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information," *CoRR*, vol. abs/2003.03308, 2020. [Online]. Available: https://arxiv.org/abs/2003.03308

[10] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6193–6206, 2017.

[11] O. Shmuel, A. Cohen, and O. Gurewitz, "Multi-antenna jamming in covert communication," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4644–4658, 2021.

[12] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Covert communication via non-causal cribbing from a cooperative jammer," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 202–207.

[13] C. N. Gagatsos, M. S. Bullock, and B. A. Bash, "Covert capacity of bosonic channels," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 2, pp. 555–567, 2020.

[14] A. Cox, Q. Zhuang, C. N. Gagatsos, B. Bash, and S. Guha, "Transceiver designs approaching the entanglement-assisted communication capacity," *Phys. Rev. Appl.*, vol. 19, p. 064015, Jun 2023.

[15] E. Zlotnick, B. A. Bash, and U. Pereg, "Entanglement-assisted covert communication via qubit depolarizing channels," *IEEE Transactions on Information Theory*, vol. 71, no. 5, pp. 3693–3706, 2025.

[16] S.-Y. Wang, S.-J. Su, and M. R. Bloch, "Resource-efficient entanglement-assisted covert communications over bosonic channels," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 3106–3111.

[17] K. S. K. Arumugam and M. R. Bloch, "Covert communication over a k -user multiple-access channel," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7020–7044, 2019.

[18] K.-H. Cho and S.-H. Lee, "Treating interference as noise is optimal for covert communication over interference channels," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 322–332, 2021.

[19] A. Bounhar, M. Sarkiss, and M. Wigger, "Whispering secrets in a crowd: Leveraging non-covert users for covert communications," 2024. [Online]. Available: https://arxiv.org/abs/2408.12962

[20] K. S. Kumar Arumugam and M. R. Bloch, "Embedding covert information in broadcast communications," *IEEE Trans. Inf. Forens. and Sec.*, vol. 14, no. 10, pp. 2787–2801, 2019.

[21] D. Kibloff, S. M. Perlaza, and L. Wang, "Embedding covert information on a given broadcast code," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 2169–2173.

[22] V. Y. F. Tan and S.-H. Lee, "Time-division is optimal for covert communication over some broadcast channels," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1377–1389, 2019.

[23] K.-H. Cho and S.-H. Lee, "Treating interference as noise is optimal for covert communication over interference channels," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 322–332, 2021.