Semantic-Security Capacity for Wiretap Channels of Type II

Ziv Goldfeld, *Student Member, IEEE*, Paul Cuff, *Member, IEEE*, and Haim H. Permuter, *Senior Member, IEEE*

Abstract—The secrecy capacity of the type II wiretap channel (WTC II) with a noisy main channel is currently an open problem. Herein its secrecy-capacity is derived and shown to be equal to its semantic-security (SS) capacity. In this setting, the legitimate users communicate via a discrete-memoryless (DM) channel in the presence of an eavesdropper that has perfect access to a subset of its choosing of the transmitted symbols, constrained to a fixed fraction of the blocklength. The secrecy criterion is achieved simultaneously for all possible eavesdropper subset choices. The SS criterion demands negligible mutual information between the message and the eavesdropper's observations even when maximized over all message distributions. A key tool for the achievability proof is a novel and stronger version of Wyner's soft covering lemma. Specifically, a random codebook is shown to achieve the soft-covering phenomenon with high probability. The probability of failure is doubly exponentially small in the blocklength. Since the combined number of messages and subsets grows only exponentially with the blocklength, SS for the WTC II is established by using the union bound and invoking the stronger soft-covering lemma. The direct proof shows that rates up to the weak-secrecy capacity of the classic WTC with a DM erasure channel (EC) to the eavesdropper are achievable. The converse follows by establishing the capacity of this DM wiretap EC as an upper bound for the WTC II. From a broader perspective, the stronger soft-covering lemma constitutes a tool for showing the existence of codebooks that satisfy exponentially many constraints, a beneficial ability for many other applications in information theoretic security.

Index Terms—Erasure wiretap channel, information theoretic security, semantic-security, soft-covering lemma, wiretap channel of type II .

I. INTRODUCTION

INFORMATION theoretic security has adopted the weak-secrecy and the strong-secrecy metrics as a standard for measuring security. Respectively, weak-secrecy and strong-secrecy refer to the normalized and unnormalized mutual

information between the secret message and the channel symbol string observed by the eavesdropper. However, recent work argues that, from a cryptographic point of view, both these metrics are insufficient to provide security of applications [1], [2]. Their main drawback lies in the assumption that the message is random and uniformly distributed, as real-life messages are neither (messages may be files, votes or any type of structured data, often with low entropy). Semantic-security (SS) [3], [4] is a cryptographic gold standard that was proposed in [2] as an adequate alternative and shown to be equivalent to a vanishing unnormalized mutual information for all message distributions. Adopting SS as our secrecy measure, we establish the SS-capacity of the wiretap channel of type II (WTC II) with an noisy main channel, for which even the secrecy-capacity was an open problem until now. On top of that, the SS-capacity and the weak-secrecy-capacity are shown to coincide.

Secret communication over noisy channels dates back to Wyner who introduced the degraded wiretap channel (WTC) and derived its weak-secrecy-capacity [5]. Csiszár and Körner extended Wyner's result to the non-degraded WTC [6], which is henceforth referred to as the WTC I. A special instance of the WTC I is when the eavesdropper's observation is an outcome of a discrete-memoryless (DM) erasure channel (EC), which essentially means that he observes a subset of the transmitted symbols which is chosen at random by nature. The WTC II was proposed by Ozarow and Wyner [7] as a generalization of this instance, where a more powerful eavesdropper selects which subset to observe and security must hold versus all possible subset choices. For this reason, the main challenge in establishing security for the WTC II boils down to finding a single sequence of codes that work well for each of the exponentially many subsets the eavesdropper may choose. In [7], the authors overcome this difficulty when the main channel is *noiseless* by relying on a unique randomized coset coding scheme in the proof of achievability. The derived rate-equivocation region was also shown to be tight, which solved the noiseless main channel scenario. The WTC II with a general (i.e., possibly *noisy*) DM main channel, however, remained an open problem ever since.

A recent endeavor at the optimal secrecy rate of the WTC II with a noisy main channel was presented in [8] (see also [9]–[12] for related work). Requiring a vanishing *average* error probability and security with respect to the *weak-secrecy* metric (namely, while assuming a uniformly distributed message and a normalized mutual information), the authors of [8] extended the coset coding scheme from [7] to obtain an inner bound on the rate-equivocation region. An outer bound was also established by assuming that the

subset the eavesdropper chooses to observe is revealed to all parties (i.e., to the legitimate users). Specializing these bounds to the maximal equivocation results in an inner and an outer bound on the weak-secrecy-capacity of a general WTC II; these bounds do not match.

In this work, we strengthen both the reliability and the security criteria, and derive the *SS-capacity* of the WTC II with a noisy main channel under a vanishing *maximal* error probability requirement. In the heart of the proof stands a stronger version of the soft-covering lemma which is key for the security analysis. Wyner's original soft-covering lemma [13, Th. 6.3] is a valuable tool for achievability proofs of information theoretic security [14]–[17], resolvability [18], channel synthesis [19], and source coding [20] (see also references therein). The result herein sharpens the claim of soft-covering by moving away from an expected value analysis. Instead, we show that a random codebook achieves the soft-covering phenomenon with high probability. The probability of failure is doubly-exponentially small in the blocklength, enabling more powerful applications through the union bound. Specifically, the lemma lets one prove the existence of code-books that satisfy exponentially many secrecy-related constraints, which, in turn, resolves the difficulty in the security analysis for the WTC II.

As a simple preliminary application of the stronger soft-covering lemma, we derive the SS-capacity of the DM-WTC I under a maximal error probability requirement. In [21], this result was established in terms of source universal coding based on the expurgation technique (e.g., cf. [22, Th. 7.7.1]) for the broadcast channel with confidential messages [6], which subsumes the WTC I as a special case. Efficient code constructions with polynomial complexity that achieve the SS-capacity under an average error probability constraint were presented in [2] for the DM scenario and in [23] for the Gaussian case, while [24] derived the Gaussian SS-capacity under a maximal error probability constraint. Complexity not being in the scope of this work, we focus on the fundamental limits of semantically-secure communication and give an alternative proof of the WTC I SS-capacity based on the stronger soft-covering lemma and classic wiretap codes. Since the number of secret messages is only exponentially large, the double-exponential decay the lemma provides ensures SS with arbitrarily high probability. In other words, even though a codebook that satisfies exponentially many constraints related to soft-covering is required, the union bound yields that such a codebook exists. This code is then amended to be reliable with respect to the maximal error probability by relying on the aforementioned expurgation technique.

Somewhat surprisingly, our optimal code construction for the WTC II is just the same. Here, SS involves a vanishing unnormalized mutual information (between the message and the eavesdropper's observation), when maximized over all message distributions and eavesdropper's subset choices. However, noting that their combined number grows only exponentially with the blocklenght, the stronger soft-covering lemma is still sharp enough to imply that the probability of an insecure random wiretap code is doubly-exponentially small. As for the WTC I, reliability is upgraded to account for maximal

error probability using expurgation. The direct proof shows that any rate up to the weak-secrecy-capacity of the WTC I with a DM-EC[1] to the eavesdropper, is achievable. The converse follows by showing that the weak-secrecy-capacity of this WTC I upper bounds the SS-capacity of the WTC II. An important consequence of the WTC II SS-capacity proof is that Wyner's wiretap codes for the erasure WTC I, are optimal. The binary version of these codes is, in fact, one of the few examples for which there are explicit constructions of practical secure encoders and decoders with optimal performance [25], [26].

This paper is organized as follows. Section II provides definitions and basic properties. In Section III we state the stronger soft-covering lemma and provide its proof. Section IV describes the WTC I and gives an alternative stronger soft-covering lemma based derivation of its SS-capacity. In Section V we define the WTC II, state its SS-capacity and prove the result. Finally, Section VI summarizes the main achievements and insights of this work.

II. NOTATIONS AND PRELIMINARIES

We use the following notations. Given two real numbers a, b, we denote by $[a : b]$ the set of integers $\{n \in \mathbb{N} | \lceil a \rceil \leq n \leq \lfloor b \rfloor\}$. We define $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$. Calligraphic letters denote sets, e.g., $\mathcal{X}$, the complement of $\mathcal{X}$ is denoted by $\mathcal{X}^c$, while $|\mathcal{X}|$ stands for its cardinality. $\mathcal{X}^n$ denoted the n-fold Cartesian product of $\mathcal{X}$. An element of $\mathcal{X}^n$ is denoted by $x^n = (x_1, x_2, \ldots, x_n)$; whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., $\mathbf{x}$. For any $\mathcal{S} \subseteq [1 : n]$, we use $\mathbf{x}^{\mathcal{S}} = (x_i)_{i \in \mathcal{S}}$ to denote the substring of x^n defined by $\mathcal{S}$, with respect to the natural ordering of $\mathcal{S}$. For instance, if $\mathcal{S} = [i : j]$, where $1 \leq i < j \leq n$, then $\mathbf{x}^{\mathcal{S}} = (x_i, x_{i+1}, \ldots, x_j)$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra and $\mathbb{P}$ is the probability measure. Random variables over $(\Omega, \mathcal{F}, \mathbb{P})$ are denoted by uppercase letters, e.g., X, with similar conventions for random vectors. The probability of an event $\mathcal{A} \in \mathcal{F}$ is denoted by $\mathbb{P}(\mathcal{A})$, while $\mathbb{P}(\mathcal{A}|\mathcal{B})$ denotes conditional probability of $\mathcal{A}$ given $\mathcal{B}$. We use $\mathbb{1}_{\mathcal{A}}$ to denote the indicator function of $\mathcal{A}$. The set of all probability mass functions (PMFs) on a finite set Ω is denoted by $\mathcal{P}(\Omega)$. PMFs are denoted by capital letters, such as P or Q, with a subscript that identifies the random variable and its possible conditioning. For example, for a discrete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two correlated random variables X and Y over that space, we use P_X, $P_{X,Y}$ and $P_{X|Y}$ to denote, respectively, the marginal PMF of X, the joint PMF of (X, Y) and the conditional PMF of X given Y. In particular, $P_{X|Y}$ represents the stochastic matrix whose elements are given by $P_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y)$. We omit subscripts if the arguments of the PMF are lowercase versions of the random variables. The support of a PMF P and the expectation of a random variable X are denoted by $\mathrm{supp}(P)$ and $\mathbb{E}[X]$, respectively.

---

[1]the erasure probability corresponds to the portion of symbols the eavesdropper in the WTC II does not intercept.
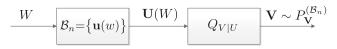
Fig. 1. Coding problem with the goal of making $P_\mathbf{V}^{(\mathcal{B}_n)} \approx Q_V^n$.

For a discrete measurable space $(\Omega, \mathcal{F})$, a PMF $Q \in \mathcal{P}(\Omega)$ gives rise to a probability measure on $(\Omega, \mathcal{F})$, which we denote by $\mathbb{P}_Q$; accordingly, $\mathbb{P}_Q(\mathcal{A}) = \sum_{\omega \in \mathcal{A}} Q(\omega)$, for every $\mathcal{A} \in \mathcal{F}$. We use $\mathbb{E}_Q$ to denote an expectation taken with respect to $\mathbb{P}_Q$. For a random variable X, we sometimes write $\mathbb{E}_X$ to emphasize that the expectation is taken with respect to P_X. For a sequence of random variable X^n, if the entries of X^n are drawn in an independent and identically distributed (i.i.d.) manner according to P_X, then for every $\mathbf{x} \in \mathcal{X}^n$ we have $P_{X^n}(\mathbf{x}) = \prod_{i=1}^n P_X(x_i)$ and we write $P_{X^n}(\mathbf{x}) = P_X^n(\mathbf{x})$. Similarly, if for every $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ we have $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, then we write $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = P_{Y|X}^n(\mathbf{y}|\mathbf{x})$. The conditional product PMF $P_{Y|X}^n$ given a specific sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted by $P_{Y|X=\mathbf{x}}^n$.

The empirical PMF $\nu_\mathbf{x}$ of a sequence $\mathbf{x} \in \mathcal{X}^n$ is

$$\nu_\mathbf{x}(x) \triangleq \frac{N(x|\mathbf{x})}{n}, \tag{1}$$

where $N(x|\mathbf{x}) = \sum_{i=1}^n \mathbb{1}_{\{x_i = x\}}$. We use $\mathcal{T}_\epsilon^n(P_X)$ to denote the set of letter-typical sequences of length n with respect to the PMF $P_X \in \mathcal{P}(\mathcal{X})$ and the non-negative number ϵ [27, Ch. 3], i.e., we have

$$\mathcal{T}_\epsilon^n(P_X) = \left\{ \mathbf{x} \in \mathcal{X}^n \middle| |\nu_\mathbf{x}(x) - P_X(x)| \leq \epsilon P_X(x), \forall x \in \mathcal{X} \right\}. \tag{2}$$

The relative entropy between two probability measures P and Q on the same σ-algebra $\mathcal{F}$ of subsets of the sample space $\mathcal{X}$, with $P \ll Q$ (i.e., P is absolutely continuous with respect to Q) is

$$D(P||Q) = \int_\mathcal{X} dP \log\left(\frac{dP}{dQ}\right), \tag{3}$$

where $\frac{dP}{dQ}$ denotes the Radon-Nikodym derivative between P and Q. If the sample space $\mathcal{X}$ is countable, (3) reduces to

$$D(P||Q) = \sum_{x \in \mathrm{supp}(P)} P(x) \log\left(\frac{P(x)}{Q(x)}\right). \tag{4}$$

III. THE STRONGER SOFT-COVERING LEMMA

Wyner's soft-covering lemma [13, Th. 6.3] states that the distribution induced by selecting a u-sequence at random from an appropriately chosen set $\mathcal{B}_n$ and passing it through a memoryless channel $Q_{V|U}^n$, results in a good approximation of Q_V^n in the limit of large n, as long as the set is of size $|\mathcal{B}_n| = 2^{nR}$, where $R > I(U; V)$ (Fig. 1). In fact, the set can be chosen quite carelessly - by random codebook construction, drawing each sequence independently from the distribution Q_U^n.

The soft-covering lemmas in the literature use a distance metric on distributions (commonly total variation or relative entropy) and claim that the distance between the induced

distribution $P_\mathbf{V}^{(\mathcal{B}_n)}$ and the desired distribution Q_V^n vanishes in expectation over the random selection of the set.[2] In the literature, [18] studies the fundamental limits of soft-covering as "resolvability", [28] provides rates of exponential convergence, [19] improves the exponents and extends the framework, [29] and [30, Ch. 16] refer to soft-covering simply as "covering" in the quantum context, [31] refers to it as a "sampling lemma" and points out that it holds for the stronger metric of relative entropy, and [32] gives a recent direct proof of the relative entropy result.

Here we give a stronger claim. With high probability with respect to the set construction, the distance vanishes exponentially quickly with the blocklength n. The negligible probability of the random set not producing this desired result is doubly-exponentially small.

Let $\mathcal{W} = \begin{bmatrix} 1 : 2^{nR} \end{bmatrix}$ and $\mathbb{B}_n = \{\mathbf{U}(w)\}_{w \in \mathcal{W}}$ be a set of random vectors that are i.i.d. according to Q_U^n. We refer to $\mathbb{B}_n$ as the random codebook. Let $\mathcal{B}_n = \{\mathbf{u}(w, \mathcal{B}_n)\}_{w \in \mathcal{W}}$ denote a realization of $\mathbb{B}_n$. For every fixed $\mathcal{B}_n$, the induced distribution is:

$$P_\mathbf{V}^{(\mathcal{B}_n)}(\mathbf{v}) = 2^{-nR} \sum_{w \in \mathcal{W}} Q_{V|U}^n\left(\mathbf{v}|\mathbf{u}(w, \mathcal{B}_n)\right). \tag{5}$$

Lemma 1 (Stronger Soft-Covering Lemma): For any Q_U, $Q_{V|U}$, and $R > I(U; V)$, where $|\mathcal{V}| < \infty$, there exist $\gamma_1, \gamma_2 > 0$, such that for n large enough

$$\mathbb{P}\left(D\left(P_\mathbf{V}^{(\mathbb{B}_n)}\middle|\middle|Q_V^n\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}. \tag{6}$$

More precisely, for any $n \in \mathbb{N}$ and $\delta \in (0, R - I(U; V))$

$$\mathbb{P}\left(D\left(P_\mathbf{V}^{(\mathbb{B}_n)}\middle|\middle|Q_V^n\right) > c_\delta n 2^{-n\gamma_\delta}\right) \leq \left(1 + |\mathcal{V}|^n\right) e^{-\frac{1}{3}2^{n\delta}}, \tag{7}$$

where

$$\gamma_\delta = \sup_{\alpha > 1} \frac{\alpha - 1}{2\alpha - 1}\left(R - \delta - d_\alpha(Q_{U,V}, Q_U Q_V)\right), \tag{8a}$$

$$c_\delta = 3 \log e + 2\gamma_\delta \log 2 + 2 \log\left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)}\right), \tag{8b}$$

and $d_\alpha(\Gamma, \Pi) = \frac{1}{\alpha - 1} \log_2 \int d\Gamma \left(\frac{d\Pi}{d\Gamma}\right)^{1-\alpha}$ is the Rényi divergence of order α.

Remark 1: The inequality (7) is trivially true for δ outside of the expressed range.

The important quantity in the lemma above is γ_δ, which is the exponent that soft-covering achieves. We see in (7) that the double-exponential convergence of probability occurs with exponent $\delta > 0$. Thus, the best soft-covering exponent that the lemma achieves with confidence, over all $\delta > 0$, is

$$\gamma^* = \sup_{\delta > 0} \gamma_\delta = \gamma_0 = \sup_{\alpha > 1} \frac{\alpha - 1}{2\alpha - 1}\left(R - d_\alpha(Q_{U,V}, Q_U Q_V)\right). \tag{9}$$

The double-exponential confidence rate δ acts as a reduction in codebook rate R in the definition of γ_δ. Consequently, $\gamma_\delta = 0$ for $\delta \geq R - I(U; V)$.

---

[2]Many of the theorems only claim existence of a good codebook, but all of the proofs use expected value to establish existence.

Remark 2 (Total Variation Exponent of Decay): The stronger soft-covering lemma can be reproduced while replacing the relative divergence with total variation [33]. Although, relative entropy can be used to bound total variation via Pinsker's inequality, this approach causes a loss of a factor of 2 in the exponent of decay. Alternatively, the proof of Lemma 1 can be modified to produce the bound on the total variation instead of the relative entropy. This direct method keeps the error exponents the same for the total variation case as it is for relative entropy.

Before proving Lemma 1, we note that the name 'stronger soft-covering lemma' is justified because (6) implies that the expectation of the relative entropy over the ensemble of codebooks decays exponentially fast (i.e., Wyner's notion of soft-covering). This is stated in the following lemma and proven in Appendix A.

Lemma 2 (Stronger than Wyner's Soft-Covering Lemma): Let $\gamma_1, \gamma_2 > 0$ be such that (6) holds for n large enough, then for every such n,

$$\mathbb{E}D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)}\middle\|Q_V^n\right) \le e^{-n\gamma_1} + n\log\left(\frac{1}{\mu_V}\right)e^{-e^{n\gamma_2}}, \qquad (10)$$

where $\mu_V = \min_{v\in\mathrm{supp}(Q_V)} Q_V(v) > 0$.

Proof of Lemma 1: We state the proof in terms of arbitrary distributions (not necessarily discrete). When needed, we will specialize to the case that $\mathcal{V}$ is finite. For any fixed codebook $\mathcal{B}_n$, let the Radon-Nikodym derivative between the induced and desired distributions be denoted as

$$\Delta_{\mathcal{B}_n}(\mathbf{v}) \triangleq \frac{dP_{\mathbf{V}}^{(\mathcal{B}_n)}}{dQ_V^n}(\mathbf{v}). \qquad (11)$$

In the discrete case, this is just a ratio of PMFs. Accordingly, the relative entropy of interest, which is a function of the codebook $\mathcal{B}_n$, is given by

$$D\left(P_{\mathbf{V}}^{(\mathcal{B}_n)}\middle\|Q_V^n\right) = \int dP_{\mathbf{V}}^{(\mathcal{B}_n)} \log \Delta_{\mathcal{B}_n}. \qquad (12)$$

To describe the jointly-typical set over u- and v-sequences, we first define information density $i_{Q_{U,V}}$, which is a function on the space $\mathcal{U} \times \mathcal{V}$ specified by

$$i_{Q_{U,V}}(u, v) \triangleq \log\left(\frac{dQ_{V|U=u}}{dQ_V}(v)\right). \qquad (13)$$

In (13), the argument of the logarithm is the Radon-Nikodym derivative between $Q_{V|U=u}$ and Q_V. Let $\epsilon \ge 0$ be arbitrary, to be determined later, and define

$$\mathcal{A}_\epsilon \triangleq \left\{(\mathbf{u}, \mathbf{v}) \in \mathcal{U}^n \times \mathcal{V}^n \middle| \frac{1}{n}i_{Q_{U,V}^n}(\mathbf{u}, \mathbf{v}) < I(U;V) + \epsilon\right\}, \qquad (14)$$

and note that

$$i_{Q_{U,V}^n}(\mathbf{u}, \mathbf{v}) = \sum_{t=1}^n i_{Q_{U,V}}(u_t, v_t). \qquad (15)$$

We split $P_{\mathbf{V}}^{(\mathcal{B}_n)}$ into two parts, making use of the indicator function. For every $\mathbf{v} \in \mathcal{V}^n$, define

$$P_{\mathcal{B}_n,1}(\mathbf{v}) \triangleq 2^{-nR} \sum_{w\in\mathcal{W}} Q_{V|U}^n\left(\mathbf{v}\middle|\mathbf{u}(w,\mathcal{B}_n)\right)\mathbb{1}_{\left\{\left(\mathbf{u}(w,\mathcal{B}_n),\mathbf{v}\right)\in\mathcal{A}_\epsilon\right\}}, \qquad (16a)$$

$$P_{\mathcal{B}_n,2}(\mathbf{v}) \triangleq 2^{-nR} \sum_{w\in\mathcal{W}} Q_{V|U}^n\left(\mathbf{v}\middle|\mathbf{u}(w,\mathcal{B}_n)\right)\mathbb{1}_{\left\{\left(\mathbf{u}(w,\mathcal{B}_n),\mathbf{v}\right)\notin\mathcal{A}_\epsilon\right\}}. \qquad (16b)$$

The measures $P_{\mathcal{B}_n,1}$ and $P_{\mathcal{B}_n,2}$ on the space $\mathcal{V}^n$ are not probability measures, but $P_{\mathcal{B}_n,1} + P_{\mathcal{B}_n,2} = P_{\mathbf{V}}^{(\mathcal{B}_n)}$ for each codebook $\mathcal{B}_n$. We also split $\Delta_{\mathcal{B}_n}$ into two parts. Namely, for every $\mathbf{v} \in \mathcal{V}^n$, we set

$$\Delta_{\mathcal{B}_n,1}(\mathbf{v}) \triangleq \frac{dP_{\mathcal{B}_n,1}}{dQ_V^n}(\mathbf{v}) \qquad (17a)$$

$$\Delta_{\mathcal{B}_n,2}(\mathbf{v}) \triangleq \frac{dP_{\mathcal{B}_n,2}}{dQ_V^n}(\mathbf{v}). \qquad (17b)$$

With respect to the above definitions, Lemma 3 states an upper bound on the relative entropy of interest.

Lemma 3: For every fixed codebook $\mathcal{B}_n$, we have

$$D\left(P_{\mathbf{V}}^{(\mathcal{B}_n)}\middle\|Q_V^n\right) \le h\left(\int dP_{\mathcal{B}_n,1}\right)$$
$$+ \int dP_{\mathcal{B}_n,1}\log\Delta_{\mathcal{B}_n,1} + \int dP_{\mathcal{B}_n,2}\log\Delta_{\mathcal{B}_n,2}, \qquad (18)$$

where $h(\cdot)$ is the binary entropy function.

The proof is relegated to Appendix B. Based on Lemma 3, if the relative entropy of interest does not decay exponentially fast, then the same is true for the terms on the right-hand side (RHS) of (18). Therefore, to establish Lemma 1, its suffices to show that the probability (with respect to a random codebook) of the RHS not vanishing exponentially fast to 0 as $n \to \infty$, is double-exponentially small.

Notice that $P_{\mathcal{B}_n,1}$ usually contains almost all of the probability. That is, for any fixed $\mathcal{B}_n$, we have

$$\int dP_{\mathcal{B}_n,2}$$
$$= 1 - \int dP_{\mathcal{B}_n,1}$$
$$= \sum_{w\in\mathcal{W}} 2^{-nR} \mathbb{P}_{Q_{V|U}^n}\left(\left(\mathbf{u}(w,\mathcal{B}_n), \mathbf{V}\right) \notin \mathcal{A}_\epsilon \middle| \mathbf{U} = \mathbf{u}(w,\mathcal{B}_n)\right). \qquad (19)$$

For a random codebook, (19) becomes

$$\int dP_{\mathbb{B}_n,2}$$
$$= \sum_{w\in\mathcal{W}} 2^{-nR} \mathbb{P}_{Q_{V|U}^n}\left(\left(\mathbf{U}(w,\mathbb{B}_n), \mathbf{V}\right) \notin \mathcal{A}_\epsilon \middle| \mathbf{U} = \mathbf{U}(w,\mathbb{B}_n)\right). \qquad (20)$$

The RHS of (20) is an average of exponentially many i.i.d. random variables bounded between 0 and 1. Furthermore,

the expected value of each one is the exponentially small probability of correlated sequences being atypical:

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{B}_n} \mathbb{P}_{Q^n_{V|U}} \Big(\big(\mathbf{U}(w, \mathbb{B}_n), \mathbf{V} \big) \notin \mathcal{A}_\epsilon \Big| \mathbf{U} = \mathbf{U}(w, \mathbb{B}_n) \Big) \\
&\quad = \mathbb{P}_{Q^n_{U,V}} \Big((\mathbf{U}, \mathbf{V}) \notin \mathcal{A}_\epsilon \Big) \\
&\quad = \mathbb{P}_{Q^n_{U,V}} \left(\sum_{t=1}^n i_{Q_{U,V}}(U_t, V_t) \geq n \big(I(U;V) + \epsilon \big) \right) \\
&\quad \overset{(a)}{=} \mathbb{P}_{Q^n_{U,V}} \left(2^{\lambda \sum_{t=1}^n i_{Q_{U,V}}(U_t, V_t)} \geq 2^{n\lambda(I(U;V)+\epsilon)} \right) \\
&\quad \overset{(b)}{\leq} \frac{\mathbb{E}_{Q^n_{U,V}} 2^{\lambda \sum_{t=1}^n i_{Q_{U,V}}(U_t, V_t)}}{2^{n\lambda(I(U;V)+\epsilon)}} \\
&\quad = \left(\frac{\mathbb{E}_{Q_{U,V}} 2^{\lambda i_{Q_{U,V}}(U,V)}}{2^{\lambda(I(U;V)+\epsilon)}} \right)^n \\
&\quad \overset{(c)}{=} 2^{n\lambda \left(\frac{1}{\lambda} \log_2 \mathbb{E}_{Q_{U,V}} \left[2^{\lambda i_{Q_{U,V}}(U;V)} \right] - I(U;V) - \epsilon \right)} \\
&\quad \overset{(d)}{=} 2^{n\lambda \left(d_{\lambda+1}(Q_{U,V}, Q_U Q_V) - I(U;V) - \epsilon \right)},
\end{aligned}
\tag{21}
$$

where (a) is true for any $\lambda \geq 0$, (b) is Markov's inequality, (c) follows by restricting λ to be strictly positive, while (d) is from the definition of the Rényi divergence of order $\lambda + 1$. We use units of bits for mutual information and Rényi divergence to coincide with the base two expression of rate. Now, substituting $\alpha = \lambda + 1$ into (21) gives

$$
\mathbb{E}_{\mathbb{B}_n} \mathbb{P}_{Q^n_{V|U}} \Big(\big(\mathbf{U}(w, \mathbb{B}_n), \mathbf{V} \big) \notin \mathcal{A}_\epsilon \Big| \mathbf{U} = \mathbf{U}(w, \mathbb{B}_n) \Big) \leq 2^{-n\beta_{\alpha,\epsilon}},
\tag{22a}
$$

where

$$
\beta_{\alpha,\epsilon} = (\alpha - 1) \big(I(U;V) + \epsilon - d_\alpha(Q_{U,V}, Q_U Q_V) \big), \tag{22b}
$$

for every $\alpha > 1$ and $\epsilon \geq 0$, over which we may optimize. The optimal choice of ϵ is apparent when all bounds of the proof are considered together (some yet to be derived), but the formula may seem arbitrary at the moment. Nevertheless, fix $\delta \in \big(0, R - I(U;V) \big)$, as found in the statement of the lemma, and set

$$
\epsilon_{\alpha,\delta} = \frac{\frac{1}{2}(R - \delta) + (\alpha - 1) d_\alpha(Q_{U,V}, Q_U Q_V)}{\frac{1}{2} + (\alpha - 1)} - I(U;V).
\tag{23}
$$

Substituting into $\beta_{\alpha,\epsilon}$ gives

$$
\beta_{\alpha,\delta} \triangleq \beta_{\alpha,\epsilon_{\alpha,\delta}} = \frac{\alpha - 1}{2\alpha - 1} \big(R - \delta - d_\alpha(Q_{U,V}, Q_U Q_V) \big).
\tag{24}
$$

Observe that $\epsilon_{\alpha,\delta}$ in (23) is nonnegative under the assumption that $R - \delta > I(U;V)$, because $\alpha > 1$ and $d_\alpha(Q_{U,V}, Q_U Q_V) \geq d_1(Q_{U,V}, Q_U Q_V) = I(U;V)$.

Next, we use the following version of the Chernoff bound to bound the probability of (20) not being exponentially small.

Lemma 4 (Chernoff Bound): Let $\{X_m\}_{m=1}^M$ be a collection of i.i.d. random variables with $X_m \in [0, B]$ and $\mathbb{E}X_m \leq \mu \neq 0$, *for all* $m \in [1 : M]$. *Then for any c with* $\frac{c}{\mu} \in [1, 2]$,

$$
\mathbb{P} \left(\frac{1}{M} \sum_{m=1}^M X_m \geq c \right) \leq e^{-\frac{M\mu}{3B} \left(\frac{c}{\mu} - 1 \right)^2}.
\tag{25}
$$

The proof is given in Appendix C.

Using (25) with $M = 2^{nR}$, $\mu = 2^{-n\beta_{\alpha,\delta}}$, $B = 1$, and $\frac{c}{\mu} = 2$, assures that $\int dP_{\mathcal{B}_n, 2}$ is exponentially small with probability doubly-exponentially close to 1. That is

$$
\mathbb{P} \left(\int dP_{\mathcal{B}_n, 2} \geq 2 \cdot 2^{-n\beta_{\alpha,\delta}} \right) \leq e^{-\frac{1}{3} 2^{n(R-\beta_{\alpha,\delta})}}.
\tag{26}
$$

Similarly, $\Delta_{\mathbb{B}_n, 1}$ is an average of exponentially many i.i.d. and uniformly bounded functions, each one determined by one sequence in the random codebook:

$$
\begin{aligned}
&\Delta_{\mathbb{B}_n, 1}(\mathbf{v}) \\
&\quad = \sum_{w \in \mathcal{W}} 2^{-nR} \frac{dQ^n_{V|U=\mathbf{U}(w, \mathbb{B}_n)}}{dQ^n_V}(\mathbf{v}) \mathbf{1}_{\left\{ \big(\mathbf{U}(w, \mathbb{B}_n), \mathbf{v} \big) \in \mathcal{A}_\epsilon \right\}}.
\end{aligned}
\tag{27}
$$

For every term in the average, the indicator function bounds the value to be between 0 and $2^{n(I(U;V)+\epsilon_{\alpha,\delta})}$. The expected value of each term with respect to the codebook is bounded above by one, which is observed by removing the indicator function. Therefore, the Chernoff bound assures that $\Delta_{\mathbb{B}_n, 1}$ is exponentially close to one for every $\mathbf{v} \in \mathcal{V}^n$. Setting $M = 2^{nR}$, $\mu = 1$, $B = 2^{n(I(U;V)+\epsilon_{\alpha,\delta})}$, and $\frac{c}{\mu} = 1 + 2^{-n\beta_{\alpha,\delta}}$ into (25), gives

$$
\begin{aligned}
\mathbb{P} \Big(\Delta_{\mathbb{B}_n, 1}(\mathbf{v}) \geq 1 + 2^{-n\beta_{\alpha,\delta}} \Big) &\leq e^{-\frac{1}{3} 2^{n(R-I(U;V)-\epsilon_{\alpha,\delta}-2\beta_{\alpha,\delta})}} \\
&= e^{-\frac{1}{3} 2^{n\delta}}, \quad \forall \mathbf{v} \in \mathcal{V}^n,
\end{aligned}
\tag{28}
$$

which decays doubly-exponentially fast for any $\delta > 0$.

At this point, we specialize to a finite set $\mathcal{V}$. Consequently, $\Delta_{\mathbb{B}_n, 2}$ is bounded as

$$
\Delta_{\mathbb{B}_n, 2}(\mathbf{v}) \leq \left(\max_{v \in \text{supp}(Q_V)} \frac{1}{Q_V(v)} \right)^n, \quad \forall \mathbf{v} \in \mathcal{V}^n,
\tag{29}
$$

with probability 1. Notice that the maximum is only over the support of Q_V, which makes this bound finite. The underlying reason for this restriction is that with probability one a conditional distribution is absolutely continuous with respect to its associated marginal distribution.

Having (26), (28) and (29), we can now bound the probability that the RHS of (18) is not exponentially small. Let $\mathcal{S}$ be the set of codebooks $\mathcal{B}_n$, such that all of the following are true:

$$
\int dP_{\mathcal{B}_n, 2} < 2 \cdot 2^{-n\beta_{\alpha,\delta}},
\tag{30a}
$$

$$
\Delta_{\mathcal{B}_n, 1}(\mathbf{v}) < 1 + 2^{-n\beta_{\alpha,\delta}}, \quad \forall \mathbf{v} \in \mathcal{V}^n,
\tag{30b}
$$

$$
\Delta_{\mathcal{B}_n, 2}(\mathbf{v}) \leq \left(\max_{v \in \text{supp}(Q_V)} \frac{1}{Q_V(v)} \right)^n, \quad \forall \mathbf{v} \in \mathcal{V}^n.
\tag{30c}
$$

First, we use the union bound, while taking advantage of the fact that the space $\mathcal{V}^n$ is only exponentially large, to show

that the probability of a random codebook not being in $\mathcal{S}$ is double-exponentially small:

$$
\begin{aligned}
\mathbb{P}\big(\mathbb{B}_n \notin \mathcal{S}\big) & \\
&\overset{(a)}{\leq} \mathbb{P}\left(\int dP_{\mathbb{B}_n,2} \geq 2 \cdot 2^{-n\beta_{\alpha,\delta}}\right) \\
&\quad + \sum_{\mathbf{v} \in \mathcal{V}^n} \mathbb{P}\left(\Delta_{\mathbb{B}_n,1}(\mathbf{v}) \geq 1 + 2^{-\beta_{\alpha,\delta}n}\right) \\
&\quad + \sum_{\mathbf{v} \in \mathcal{V}^n} \mathbb{P}\left(\Delta_{\mathbb{B}_n,2}(\mathbf{v}) > \left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)}\right)^n\right) \\
&\overset{(b)}{\leq} e^{-\frac{1}{3}2^{n(R-\beta_{\alpha,\delta})}} + |\mathcal{V}|^n \cdot e^{-\frac{1}{3}2^{n\delta}} \\
&\overset{(c)}{\leq} \left(1 + |\mathcal{V}|^n\right) e^{-\frac{1}{3}2^{n\delta}},
\end{aligned}
\tag{31}
$$

where (a) is the union bound, (b) uses (26), (28) and (29), while (c) follows because $\beta_{\alpha,\delta} \leq \frac{1}{2}(R-\delta)$.

Next, we claim that for every codebook in $\mathcal{S}$, the RHS of (18) is exponentially small. Let $\mathcal{B}_n \in \mathcal{S}$ and consider the following. For every $x \in [0,1]$, $h(x) \leq x \log \frac{e}{x}$, using which (30a) implies that

$$
\begin{aligned}
h\left(\int dP_{\mathcal{B}_n,1}\right) = h\left(\int dP_{\mathcal{B}_n,2}\right) & \\
&< 2\big(\log e + \beta_{\alpha,\delta} \log 2\big)n2^{-n\beta_{\alpha,\delta}}.
\end{aligned}
\tag{32}
$$

Furthermore, by (30b), we have

$$
\begin{aligned}
\int dP_{\mathcal{B}_n,1} \log \Delta_{\mathcal{B}_n,1} &< \int dP_{\mathcal{B}_n,1} \log(1 + 2^{-n\beta_{\alpha,\delta}}) \\
&\leq \log(1 + 2^{-n\beta_{\alpha,\delta}}) \\
&\overset{(a)}{\leq} 2^{-n\beta_{\alpha,\delta}} \log e,
\end{aligned}
\tag{33}
$$

where (a) follows since $\log(1+x) \leq x \log e$, for every $x > 0$. Finally, using (30c) we obtain

$$
\begin{aligned}
\int dP_{\mathcal{B}_n,2} \log \Delta_{\mathcal{B}_n,2} &\leq \int dP_{\mathcal{B}_n,2} \log \left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)}\right)^n \\
&< 2 \log \left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)}\right) n2^{-n\beta_{\alpha,\delta}}.
\end{aligned}
\tag{34}
$$

Combining (32)-(34), yields

$$
\begin{aligned}
h\left(\int dP_{\mathcal{B}_n,1}\right) &+ \int dP_{\mathcal{B}_n,1} \log \Delta_{\mathcal{B}_n,1} + \int dP_{\mathcal{B}_n,2} \log \Delta_{\mathcal{B}_n,2} \\
&< \Bigg(2\big(\log e + \beta_{\alpha,\delta} \log 2\big) + \log e \\
&\quad + 2 \log \left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)}\right)\Bigg) n2^{-n\beta_{\alpha,\delta}} \\
&\overset{(a)}{=} c_{\alpha,\delta}n2^{-n\beta_{\alpha,\delta}}
\end{aligned}
\tag{35}
$$

where (a) comes from setting

$$
c_{\alpha,\delta} \triangleq 3\log e + 2\beta_{\alpha,\delta}\log 2 + 2\log \left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)}\right).
\tag{36}
$$

This implies that for all $\alpha > 1$ and $\delta \in \big(0, R - I(U;V)\big)$,

$$
\begin{aligned}
\mathbb{P}\left(D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \big\| Q_V^n\right) \geq c_{\alpha,\delta}n2^{-n\beta_{\alpha,\delta}}\right) & \\
&\leq \mathbb{P}\Bigg(h\left(\int dP_{\mathbb{B}_n,1}\right) + \int dP_{\mathbb{B}_n,1} \log \Delta_{\mathbb{B}_n,1} \\
&\quad + \int dP_{\mathbb{B}_n,2} \log \Delta_{\mathbb{B}_n,2} \geq c_{\alpha,\delta}n2^{-n\beta_{\alpha,\delta}}\Bigg) \\
&\leq \mathbb{P}\big(\mathbb{B}_n \notin \mathcal{S}\big) \\
&\overset{(a)}{\leq} \left(1 + |\mathcal{V}|^n\right) e^{-\frac{1}{3}2^{n\delta}},
\end{aligned}
\tag{37}
$$

where (a) follows from (31). Denoting $c_\delta \triangleq \sup_{\alpha>1} c_{\alpha,\delta}$, (37) further gives

$$
\mathbb{P}\left(D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \big\| Q_V^n\right) \geq c_\delta n2^{-n\beta_{\alpha,\delta}}\right) \leq \left(1 + |\mathcal{V}|^n\right) e^{-\frac{1}{3}2^{n\delta}}.
\tag{38}
$$

Since (38) is true for all $\alpha > 1$, it must also be true, with strict inequality in the LHS, when replacing $\beta_{\alpha,\delta}$ with

$$
\gamma_\delta \triangleq \sup_{\alpha>1} \beta_{\alpha,\delta} = \sup_{\alpha>1} \frac{\alpha-1}{2\alpha-1}\big(R - \delta - d_\alpha(Q_{U,V}, Q_U Q_V)\big),
\tag{39}
$$

which is the exponential rate of convergence stated in (8a) that we derive for the strong soft-covering lemma. This establishes the statement from (7) and proves Lemma 1.

Concluding, if $R > I(U;V)$ and for any $\delta \in \big(0, R - I(U;V)\big)$, we get exponential convergence of the relative entropy at rate $O(2^{-\gamma_\delta n})$ with doubly-exponential certainty. Discarding the precise exponents of convergence and coefficients, we state that there exist $\gamma_1, \gamma_2 > 0$, such that for n large enough

$$
\mathbb{P}\left(D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \big\| Q_V^n\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}.
\tag{40}
$$

■

IV. WIRETAP CHANNEL I

As a rather simple application of stronger soft-covering lemma, we give an alternative derivation of the SS-capacity of the WTC I [2], [21], [23], [24]. Since the channel to the legitimate user is the same in both WTCs I and II, the maximal error probability analysis presented here is subsequently used to establish reliability for the WTC II.

Our direct proof relies on classic wiretap codes and SS is established using the union bound while invoking the stronger soft-covering lemma. In a wiretap code, a subcode is associated with each confidential message. To transmit a certain message, a codeword from its subcode is selected uniformly at random and transmitted over the channel. Letting these subcodes be large enough while noting that the number of confidential messages only grows exponentially with the block-length, the union bound and the double-exponential decay the lemma provides show the existence of a semantically-secure sequences of codes. Using these codes, each transmitted message induces an output PMF at the eavesdropper that appears i.i.d. and does not depend on the message.
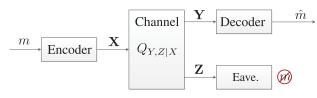
Fig. 2. The classic wiretap channel, referred to as the WTC I.

Wyner's soft-covering lemma, that is now a standard tool for achieving strong-secrecy for the WTC I, comes up short in providing SS. The classic soft-covering argument says that on average over the messages, the output at the eavesdropper will look i.i.d., provided that the size of these subcodes is large enough. This can be used to claim that the unnormalized mutual information between the message and the eavesdropper's output is small, thus ensuring strong-secrecy. However, for SS, it must be claimed that the output PMF is close the i.i.d. distribution for all messages, and there are exponentially many messages. Here is where the stronger soft-covering lemma is advantageous.

A. Problem Definition

The DM-WTC I is illustrated in Fig. 2. The sender chooses a message m from the set $[1 : 2^{nR}]$ and maps it into a sequence $\mathbf{x} \in \mathcal{X}^n$ (the mapping may be random). The sequence $\mathbf{x}$ is transmitted over the DM-WTC I with transition probability $Q_{Y,Z|X}$. The output sequences $\mathbf{y} \in \mathcal{Y}^n$ and $\mathbf{z} \in \mathcal{Z}^n$ are observed by the receiver and the eavesdropper, respectively. Based on $\mathbf{y}$, the receiver produces an estimate $\hat{m}$ of m. The eavesdropper tries to glean whatever it can about the message from $\mathbf{z}$.

Definition 1 (Code Description): An (n, R) WTC I code $\mathcal{C}_n$ has:

1) *A message set $\mathcal{M} = [1 : 2^{nR}]$.*
2) *A stochastic encoder $f : \mathcal{M} \to \mathcal{P}(\mathcal{X}^n)$.*
3) *A decoding function $\phi : \mathcal{Y}^n \to \hat{\mathcal{M}}$, where $\hat{\mathcal{M}} = \mathcal{M} \cup \{e\}$ and $e \notin \mathcal{M}$.*

For any message distribution $P_M \in \mathcal{P}(\mathcal{M})$, the joint PMF over $\mathcal{M} \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \times \hat{\mathcal{M}}$ induced by P_M and an (n, R) code $\mathcal{C}_n$ is:

$$P^{(\mathcal{C}_n)}(m, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{m}) = P_M(m) f(\mathbf{x}|m)$$
$$\times Q_{Y,Z|X}(\mathbf{y}, \mathbf{z}|\mathbf{x}) \mathbb{1}_{\{\hat{m} = \phi(\mathbf{y})\}}. \quad (41)$$

Definition 2 (Maximal Error Probability): The maximal error probability of an (n, R) WTC I code $\mathcal{C}_n$ is

$$e^\star(\mathcal{C}_n) = \max_{m \in \mathcal{M}} e_m(\mathcal{C}_n), \quad (42a)$$

where

$$e_m(\mathcal{C}_n) = \sum_{\mathbf{x} \in \mathcal{X}^n} f(\mathbf{x}|m) \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \phi(\mathbf{y}) \neq m}} Q_{Y|X}^n(\mathbf{y}|\mathbf{x}). \quad (42b)$$

Definition 3 (SS Metric): The SS metric associated to an (n, R) WTC I code $\mathcal{C}_n$ is [3]

$$\text{Sem}(\mathcal{C}_n) = \max_{P_M \in \mathcal{P}(\mathcal{M})} I_{\mathcal{C}_n}(M; \mathbf{Z}), \quad (43)$$

[3]$\text{Sem}(\mathcal{C}_n)$ is actually the mutual-information-security (MIS) metric, which is equivalent to SS by [2]. We use the representation in (43) rather than the formal definition of SS (see, e.g., [2, eq. (4)]) out of analytical convenience.

where $I_{\mathcal{C}_n}$ denotes a mutual information term that is calculated with respect to the PMF induced by $\mathcal{C}_n$ from (41).

Definition 4 (Semantically-Secure Codes): A sequence of (n, R) WTC I codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ is semantically-secure if there is a constants $\gamma > 0$ and an $n_0 \in \mathbb{N}$, such that for every $n > n_0$, $\text{Sem}(\mathcal{C}_n) \leq e^{-n\gamma}$.

Remark 3: SS requires that a single sequence of codes works well for all message PMFs. Accordingly, the mutual information term in (43) is maximized over P_M when the code $\mathcal{C}_n$ is known. In other words, although not stated explicitly, P_M is a function of $\mathcal{C}_n$.

Remark 4: By Definition 4, for a sequence of WTC I codes to be semantically-secure, the SS metric from (43) must vanish exponentially fast. This is a standard requirement in the cryptography community, commonly referred to as strong-SS (see, e.g., [2, Sec. 3.2]). The coding scheme given in the direct proof of Theorem 1 achieves this exponential decay of the SS-metric (see Section IV-C.1). An exponential decay of the strong-secrecy metric was previously observed in [21], [28], and [34].

Definition 5 (SS-Achievability): A rate $R \in \mathbb{R}_+$ is SS-achievable if there is a sequence of (n, R) WTC I semantically-secure codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ with $e^\star(\mathcal{C}_n) \to 0$ as $n \to \infty$.

Definition 6 (SS-Capacity): The SS-capacity of the WTC I, C_{Sem}, is the supremum of the set of SS-achievable rates.

B. Results

As stated in the following theorem, the SS-capacity of the WTC I under a maximal error probability constraint is the same as its weak-secrecy-capacity under an average error probability constraint.

Theorem 1 (WTC I SS-Capacity): The SS-capacity of the WTC I is

$$C_{\text{Sem}} = \max_{\substack{Q_{U,X}: \\ U - X - (Y,Z)}} \Big[I(U; Y) - I(U; Z) \Big], \quad (44)$$

and one may restrict the cardinality of U to $|\mathcal{U}| \leq |\mathcal{X}|$.

The proof of Theorem 1 is given in Section IV-C.1. Our achievability proof relies on the stronger soft-covering lemma to establish the existence of a sequence of semantically-secure codes with a vanishing average probability of error. The expurgation technique [22, Th. 7.7.1] is then used to upgrade the codes to have a vanishing maximal error probability.

Remark 5: The cardinality bound in Theorem 1 was established in [35, Th. 22.1].

Remark 6: The direct part of Theorem 1 can also be derived without using the stronger soft-covering lemma. Instead, one may invoke the codebook expurgation technique twice. By removing a certain portion of the messages, any sequence of codes that ensures strong-secrecy and a vanishing average error probability, can be upgraded to provide SS and reliability with respect to the maximal error probability with negligible rate-loss. In the original codes, the fraction of messages that induce an error probability greater than three times the average, is less than $\frac{1}{3}$. Similarly, the fraction of messages with secrecy distance greater than three times the average is

less than $\frac{1}{3}$. *Therefore, the fraction of offending messages is less than* $\frac{2}{3}$. *By removing them one obtains a new sequence of codes that is semantically-secure and has a vanishing maximal error probability. Finally, the rate of the n-th code in the new sequence is* $R - \frac{\log(3)}{n}$ *(here R stands for the rate of the original codes), and the loss is negligible for large n.*

Remark 7: The expurgation method is insufficient for establishing SS for the WTC II because the messages that need to be removed might differ from one choice of the eavesdropper's observations to the next. It also does not work in other settings such as the multiple access WTC, where expurgation is problematic in general. On the other hand, even for that setting, an achievability proof that relies on the stronger soft-covering lemma goes through by similar steps to those presented below. Thus, strong-secrecy can be upgraded to SS even in situations where vanishing average error probability cannot be upgraded to vanishing maximum error probability (via expurgation).

C. Proofs

1) Theorem 1: For the converse, let $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ be a sequence of (n, R) semantically-secure WTC I codes with $e^\star(\mathcal{C}_n) \to 0$. Since both $e^\star(\mathcal{C}_n) \to 0$ and $\mathrm{Sem}(\mathcal{C}_n) \to 0$ hold for any message distribution $P_M \in \mathcal{P}(\mathcal{M})$, in particular, they hold for a uniform P_M. The converse thus follows since C_{Sem} in (44) coincides with the secrecy-capacity of the WTC I under a vanishing average error probability criterion and the weak-secrecy constraint.

For the direct part, we first establish the achievability of (44) when $U = X$. Then, a standard channel prefixing argument extends the proof to any U with $U - X - (Y, Z)$.

Fix $\epsilon > 0$, a PMF $Q_X \in \mathcal{P}(\mathcal{X})$, and let M and W be independent random variables uniformly distributed over $\mathcal{M}$ and $\mathcal{W} \triangleq [1 : 2^{n\tilde{R}}]$, respectively. M represents the choice of the message, while W stands for the stochastic part of the encoder. Thus, we start by imposing a uniform distribution over the set of messages and use this to show the existence of a semantically-secure sequence of (n, R) codes with a vanishing *average* error probability. Afterwards, the uniform message distribution assumption is dropped using the expurgation technique [22, Th. 7.7.1], which allows upgrading reliability to achieve a vanishing *maximal* error probability, while preserving SS.

Codebook $\mathcal{B}_n$: Let $\mathbb{B}_n$ be a *random codebook* given by a collection of i.i.d. random vectors $\mathbb{B}_n = \{\mathbf{X}(m, w)\}_{(m,w) \in \mathcal{M} \times \mathcal{W}}$, each distributed according to Q_X^n. A realization of $\mathbb{B}_n$ is denoted by $\mathcal{B}_n \triangleq \{\mathbf{x}(m, w, \mathcal{B}_n)\}_{(m,w) \in \mathcal{M} \times \mathcal{W}}$, with respect to which a classic wiretap code is constructed.

Encoder f: To send $m \in \mathcal{M}$ the encoder randomly and uniformly chooses $W = w$ from $\mathcal{W}$ and transmits $\mathbf{x}(m, w, \mathcal{B}_n)$ over the WTC I.

Decoder ϕ: Upon observing $\mathbf{y} \in \mathcal{Y}^n$, the decoder searches for a unique pair $(\hat{m}, \hat{w}) \in \mathcal{M} \times \mathcal{W}$ such that

$$\big(\mathbf{x}(\hat{m}, \hat{w}, \mathcal{B}_n), \mathbf{y}\big) \in \mathcal{T}_\epsilon^n(Q_{X,Y}). \tag{45}$$

If such a unique pair is found, then set $\phi(\mathbf{y}) = \hat{m}$; otherwise, $\phi(\mathbf{y}) = e$.

The triple $(\mathcal{M}, f, \phi)$ defined with respect to the codebook $\mathcal{B}_n$ constitutes an (n, R) WTC I code $\mathcal{C}_n$. When a random codebook $\mathbb{B}_n$ is used, we denote the corresponding random code by $\mathbb{C}_n$.

Average Error Probability Analysis: By standard joint typicality arguments we show that the average error probability, when expected over the ensemble of codebooks, is arbitrarily small. For every fixed codebook $\mathcal{B}_n$ and $(\tilde{m}, \tilde{w}) \in \mathcal{M} \times \mathcal{W}$, define the event

$$\mathcal{E}(\tilde{m}, \tilde{w}, \mathcal{B}_n) = \Big\{\big(\mathbf{x}(\tilde{m}, \tilde{w}, \mathcal{B}_n), \mathbf{Y}\big) \in \mathcal{T}_\epsilon^n(Q_{X,Y})\Big\}, \tag{46}$$

where $\mathbf{Y} \sim Q_{Y|X=\mathbf{x}(\tilde{m}, \tilde{w}, \mathcal{B}_n)}^n$ is the random sequence observed at the receiver when the transmitted sends $(\tilde{m}, \tilde{w})$. We have

$$\mathbb{E}_{\mathbb{C}_n} \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_m(\mathbb{C}_n)$$

$$= \mathbb{E}_{\mathbb{C}_n} \mathbb{P}_{\mathbb{C}_n}(\hat{M} \neq M)$$

$$\leq \mathbb{E}_{\mathbb{C}_n} \mathbb{P}_{\mathbb{C}_n}\big((\hat{M}, \hat{W}) \neq (M, W)\big)$$

$$\overset{(a)}{=} \mathbb{E}_{\mathbb{C}_n} \mathbb{P}_{\mathbb{C}_n}\big((\hat{M}, \hat{W}) \neq (1, 1) \big| M = 1, W = 1\big)$$

$$\overset{(b)}{=} \mathbb{E}_{\mathbb{B}_n} \mathbb{P}\left(\mathcal{E}(1, 1, \mathbb{B}_n)^c \cup \left\{\bigcup_{(\tilde{m}, \tilde{w}) \neq (1,1)} \mathcal{E}(\tilde{m}, \tilde{w}, \mathbb{B}_n)\right\} \bigg| \mathbb{B}_n\right)$$

$$\overset{(c)}{\leq} \underbrace{\mathbb{P}_{Q_{X,Y}^n}\Big((\mathbf{X}, \mathbf{Y}) \in \mathcal{T}_\epsilon^n(Q_{X,Y})\Big)}_{P_1}$$

$$+ \underbrace{\sum_{(\tilde{m}, \tilde{w}) \neq (1,1)} \mathbb{P}_{Q_X^n \times Q_Y^n}\Big((\mathbf{X}, \mathbf{Y}) \in \mathcal{T}_\epsilon^n(Q_{X,Y})\Big)}_{P_2}, \tag{47}$$

where (a) uses the symmetry of the codebook construction with respect to m and w, (b) follows by the decoding rule, while (c) takes the expectation over the ensemble of codebooks and uses the union bound.

By the law of large numbers $P_1 \to 0$ as $n \to \infty$, while $P_2 \to 0$ as n grows provided that[4]

$$R + \tilde{R} < I(X; Y). \tag{48}$$

Thus, we have

$$\mathbb{E}_{\mathbb{C}_n} \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_m(\mathbb{C}_n) \xrightarrow[n \to \infty]{} 0. \tag{49}$$

Security Analysis: For any fixed $\mathcal{B}_n$ (which, in turn, fixes $\mathcal{C}_n$), we denote by $P_{M,\mathbf{Z}}^{(\mathcal{C}_n)}$ the joint distribution of M and $\mathbf{Z}$ induced by the code $\mathcal{C}_n$ (see (41)). For any $\mathcal{B}_n$, we first have

$$\max_{P_M \in \mathcal{P}(\mathcal{M})} I_{\mathcal{C}_n}(M; \mathbf{Z})$$

$$\overset{(a)}{=} \max_{P_M \in \mathcal{P}(\mathcal{M})} D\Big(P_{\mathbf{Z}|M}^{(\mathcal{C}_n)} \Big\| P_{\mathbf{Z}}^{(\mathcal{C}_n)} \Big| P_M\Big)$$

$$\overset{(b)}{\leq} \max_{P_M \in \mathcal{P}(\mathcal{M})} D\Big(P_{\mathbf{Z}|M}^{(\mathcal{C}_n)} \Big\| Q_Z^n \Big| P_M\Big)$$

$$= \max_{P_M \in \mathcal{P}(\mathcal{M})} \sum_{m \in \mathcal{M}} P(m) D\Big(P_{\mathbf{Z}|M=m}^{(\mathcal{C}_n)} \Big\| Q_Z^n\Big)$$

$$\leq \max_{P_M \in \mathcal{P}(\mathcal{M})} \sum_{m \in \mathcal{M}} P(m) \max_{\tilde{m} \in \mathcal{M}} D\Big(P_{\mathbf{Z}|M=\tilde{m}}^{(\mathcal{C}_n)} \Big\| Q_Z^n\Big)$$

$$= \max_{m \in \mathcal{M}} D\Big(P_{\mathbf{Z}|M=m}^{(\mathcal{C}_n)} \Big\| Q_Z^n\Big), \tag{50}$$

---

[4]All subsequent mutual information terms in the proof are calculated with respect to $Q_{U,X} Q_{Y,Z|X}$ or its marginals.

where (a) uses the relative entropy chain rule, while is because for any $P_M \in \mathcal{P}(\mathcal{M})$, we have

$$D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n)}\middle\|P_{\mathbf{Z}}^{(\mathcal{C}_n)}\middle|P_M\right)$$

$$= \sum_{m \in \mathcal{M}} P(m) \sum_{\mathbf{z} \in \mathcal{Z}^n} P^{(\mathcal{C}_n)}(\mathbf{z}|m) \log\left(\frac{P^{(\mathcal{C}_n)}(\mathbf{z}|m)}{P^{(\mathcal{C}_n)}(\mathbf{z})} \cdot \frac{Q_Z^n(\mathbf{z})}{Q_Z^n(\mathbf{z})}\right)$$

$$= D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n)}\middle\|Q_Z^n\middle|P_M\right)$$

$$- \sum_{m \in \mathcal{M}} P(m) \sum_{\mathbf{z} \in \mathcal{Z}^n} P^{(\mathcal{C}_n)}(\mathbf{z}|m) \log\left(\frac{P^{(\mathcal{C}_n)}(\mathbf{z})}{Q_Z^n(\mathbf{z})}\right)$$

$$= D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n)}\middle\|Q_Z^n\middle|P_M\right) - D\left(P_{\mathbf{Z}}^{(\mathcal{C}_n)}\middle\|Q_Z^n\right)$$

$$\leq D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n)}\middle\|Q_Z^n\middle|P_M\right). \tag{51}$$

Now, let $\tilde{\gamma}$ be an arbitrary positive real number to be determined later and consider the following probability.

$$\mathbb{P}\left(\left\{\mathrm{Sem}(\mathbb{C}_n) \leq e^{-n\tilde{\gamma}}\right\}^c\right)$$

$$\overset{(a)}{\leq} \mathbb{P}\left(\left\{\max_{m \in \mathcal{M}} D\left(P_{\mathbf{Z}|M=m}^{(\mathbb{C}_n)}\middle\|Q_Z^n\right) \leq e^{-n\tilde{\gamma}}\right\}^c\right)$$

$$= \mathbb{P}\left(\left\{\forall m \in \mathcal{M},\ D\left(P_{\mathbf{Z}|M=m}^{(\mathbb{C}_n)}\middle\|Q_Z^n\right) \leq e^{-n\tilde{\gamma}}\right\}^c\right)$$

$$= \mathbb{P}\left(\exists m \in \mathcal{M},\ D\left(P_{\mathbf{Z}|M=m}^{(\mathbb{C}_n)}\middle\|Q_Z^n\right) > e^{-n\tilde{\gamma}}\right)$$

$$= \mathbb{P}\left(\bigcup_{m \in \mathcal{M}}\left\{D\left(P_{\mathbf{Z}|M=m}^{(\mathbb{C}_n)}\middle\|Q_Z^n\right) > e^{-n\tilde{\gamma}}\right\}\right)$$

$$\leq \sum_{m \in \mathcal{M}} \mathbb{P}\left(D\left(P_{\mathbf{Z}|M=m}^{(\mathbb{C}_n)}\middle\|Q_Z^n\right) > e^{-n\tilde{\gamma}}\right), \tag{52}$$

where (a) follows from (50) and (50).

By the stronger soft-covering lemma, if

$$\tilde{R} > I(X;Z), \tag{53}$$

then there are $\gamma_1, \gamma_2 >$ such that

$$\mathbb{P}\left(D\left(P_{\mathbf{Z}|M=m}^{(\mathbb{C}_n)}\middle\|Q_Z^n\right) > e^{-n\gamma_1}\right) \leq e^{-e^{n\gamma_2}}, \tag{54}$$

for any $m \in \mathcal{M}$ and sufficiently large n. Inserting (54) into (52) while setting $\tilde{\gamma} = \gamma_1$, we have

$$\mathbb{P}\left(\left\{\mathrm{Sem}(\mathbb{C}_n) \leq e^{-n\gamma_1}\right\}^c\right) \leq \sum_{m \in \mathcal{M}} e^{-e^{n\gamma_2}}$$

$$= 2^{nR} \cdot e^{-e^{n\gamma_2}}$$

$$\triangleq \eta_n \xrightarrow{n \to \infty} 0, \tag{55}$$

and therefore,

$$\mathbb{P}\left(\mathrm{Sem}(\mathbb{C}_n) \leq e^{-n\gamma_1}\right) \geq 1 - \eta_n \xrightarrow{n \to \infty} 1. \tag{56}$$

Inequality (56) implies that if $\tilde{R}$ satisfies (53), the probability that a randomly generated sequence of codes meets the SS criterion for large n is arbitrarily close to 1. In fact, because (55) decays so rapidly, the Borel-Cantelli lemma implies that almost every sequence of realizations of $\{\mathbb{C}_n\}_{n \in \mathbb{N}}$ is semantically-secure.

SS-Achievability: To establish the existence of a sequence of $(n, 2^{nR})$ reliable and semantically-secure codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$, we reproduce the Selection Lemma [36, Lemma 2.2].

Lemma 5 (Selection Lemma): Let $\{A_n\}_{n \in \mathbb{N}}$ be a sequence of random variables, where A_n takes values in $\mathcal{A}_n$. Let $\{f_n^{(1)}, f_n^{(2)}, \ldots, f_n^{(I)}\}_{n \in \mathbb{N}}$ be a collection of $I < \infty$ sequences of bounded functions $f_n^{(i)} : \mathcal{A}_n \to \mathbb{R}_+$, $i \in [1 : I]$. If

$$\mathbb{E} f_n^{(i)}(A_n) \xrightarrow{n \to \infty} 0, \quad \forall i \in [1 : I], \tag{57a}$$

then there exists a sequence $\{a_n\}_{n \in \mathbb{N}}$, where $a_n \in \mathcal{A}_n$ for every $n \in \mathbb{N}$, such that

$$f_n^{(i)}(a_n) \xrightarrow{n \to \infty} 0, \quad \forall i \in [1 : I]. \tag{57b}$$

For completeness, the proof of Lemma 5 is given in Appendix D. Applying Lemma 5 to the random variables $\{\mathbb{C}_n\}_{n \in \mathbb{N}}$ and the functions $\frac{1}{|\mathcal{M}|}\sum_{m \in \mathcal{M}} e_m(\mathbb{C}_n)$ and $\mathbb{1}_{\left\{\mathrm{Sem}(\mathbb{C}_n) > e^{-n\gamma_1}\right\}}$, while using (49) and (55), we have that there is a sequence of (n, R) WTC I codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$, for which

$$\frac{1}{|\mathcal{M}|}\sum_{m \in \mathcal{M}} e_m(\mathcal{C}_n) \xrightarrow{n \to \infty} 0, \tag{58a}$$

$$\mathbb{1}_{\left\{\mathrm{Sem}(\mathcal{C}_n) > e^{-n\gamma_1}\right\}} \xrightarrow{n \to \infty} 0. \tag{58b}$$

Since the indicator function in (58b) takes only the values 0 and 1, to satisfy the convergence there must exist an $n_0 \in \mathbb{N}$, such that

$$\mathbb{1}_{\left\{\mathrm{Sem}(\mathcal{C}_n) > e^{-n\gamma_1}\right\}} = 0, \quad \forall n > n_0, \tag{59}$$

and therefore,

$$\mathrm{Sem}(\mathcal{C}_n) \leq e^{-n\gamma_1}, \quad \forall n > n_0. \tag{60}$$

The final step is to amend $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ to be reliable with respect to the maximal error probability (as defined in (42a)). This is done using the expurgation technique (see, e.g., [22, Th. 7.7.1]). Namely, we discard the worst half of the codewords in each codebook $\mathcal{B}_n$. Denoting the amended sequence of codebooks by $\{\mathcal{B}_n^\star\}_{n \in \mathbb{N}}$ and their corresponding sequence of codes by $\{\mathcal{C}_n^\star\}_{n \in \mathbb{N}}$, we have

$$e^\star(\mathcal{C}_n^\star) \xrightarrow{n \to \infty} 0. \tag{61}$$

Note that in each $\mathcal{C}_n^\star$ there are 2^{nR-1} codewords, i.e., throwing out half the codewords has changed the rate from R to $R - \frac{1}{n}$, which is negligible for large n. Further note that because $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ is semantically-secure, so is $\{\mathcal{C}_n^\star\}_{n \in \mathbb{N}}$. Combining (48) with (53), we have that every

$$0 \leq R < \max_{Q_X}\left[I(X;Y) - I(X;Z)\right] \tag{62}$$

is SS-achievable.

To establish the achievability of C_{Sem} from (44), we prefix a DM-channel (DMC) $Q_{X|U}$ to the original WTC I $Q_{Y,Z|X}$ to obtain a new channel $Q_{Y,Z|U}$, where

$$Q_{Y,Z|U}^n(\mathbf{y}, \mathbf{z}|\mathbf{u}) = \sum_{\mathbf{x} \in \mathcal{X}^n} Q_{X|U}^n(\mathbf{x}|\mathbf{u}) Q_{Y,Z|X}^n(\mathbf{y}, \mathbf{z}|\mathbf{x}). \tag{63}$$
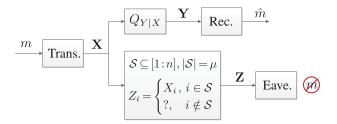
Fig. 3.   The type II wiretap channel.

Using a similar analysis as above with respect to $Q_{Y,Z|U}$, any $R \in \mathbb{R}^+$ satisfying

$$R < \max_{\substack{Q_{U,X}: \\ U-X-(Y,Z)}} \left[I(U;Y) - I(U;Z) \right] \qquad (64)$$

is achievable.

V. WIRETAP CHANNEL II

The WTC II scenario considers communication between two legitimate parties in the presence of an eavesdropper that can choose to observed any subset of the transmitted sequence, while being limited in quantity. The challenge in this setting is that the eavesdropper knows the codebook when it selects the subset to observe. Therefore, secrecy will only be achieved if it is achieved uniformly for all selections of packets, of which there are exponentially many possibilities. Furthermore, SS being our goal, secrecy must be ensured for each one of the exponentially many confidential messages. Nonetheless, as the combined number of subsets and messages grows only exponentially with the blocklenght, using the stronger soft-covering lemma we show that rates all the way up to the weak-secrecy-capacity of the DM erasure WTC I are achievable even in this more stringent setting. Then, we establish the secrecy-capacity of this WTC I as an upper bound on the considered WTC II, thus characterizing the SS-capacity of the latter.

A. Problem Definition

The WTC II is illustrated in Fig. 3. The sender chooses a message m from the set $\left[1 : 2^{nR} \right]$ and maps it into a sequence $\mathbf{x} \in \mathcal{X}^n$ (the mapping may be random). The sequence $\mathbf{x}$ is transmitted over a point-to-point DMC with transition probability $Q_{Y|X}$. Based on the received channel output sequence $\mathbf{y} \in \mathcal{Y}^n$, the receiver produces an estimate $\hat{m}$ of m. The eavesdropper noiselessly observes a subset of its choice of the n transmitted symbols. Namely, the eavesdropped chooses $\mathcal{S} \subseteq [1 : n]$, $|\mathcal{S}| = \mu \leq n$, and observes $\mathbf{z} \in \left(\mathcal{X} \cup \{?\} \right)^n$, where

$$z_i = \begin{cases} x_i, & i \in \mathcal{S} \\ ?, & i \notin \mathcal{S}. \end{cases} \qquad (65)$$

Based on $\mathbf{z}$, the eavesdropper tries to learn as much as possible about the message.

With some abuse of notation (reusing notations from Section IV-A), we introduce the following definitions.

An (n, R) WTC II code $\mathcal{C}_n$ and the corresponding maximal error probability $e^\star(\mathcal{C}_n)$ are defined similarly to Definitions 1 and 2, respectively.

Definition 7 (SS Metric): The SS metric with respect to an (n, R) WTC II code $\mathcal{C}_n$ is

$$\mathrm{Sem}_\mu(\mathcal{C}_n) = \max_{\substack{P_M \in \mathcal{P}(\mathcal{M}), \\ \mathcal{S} \subseteq [1:n]: \ |\mathcal{S}|=\mu}} I_{\mathcal{C}_n}(M; \mathbf{Z}), \qquad (66)$$

where $I_{\mathcal{C}_n}$ denotes that the mutual information term is calculated with respect to

$$P_{M,\mathbf{Z}}^{(\mathcal{C}_n, \mathcal{S})}(m, \mathbf{z}) = P(m) \sum_{\mathbf{x} \in \mathcal{X}^n} f(\mathbf{x}|m) \mathbb{1}_{\{z_i=x_i, i \in \mathcal{S}\} \cap \{z_i=?, i \notin \mathcal{S}\}}.$$

Remark 8: As explained in Remark 3, the code $\mathcal{C}_n$ is known when the mutual information term in (66) is maximized. Thus, the observed subset $\mathcal{S} \subseteq [1 : n]$ and the message PMF P_M are both functions of $\mathcal{C}_n$. Although, for the sake of simplicity, this dependence is omitted from our notations, the reader should keep in mind that a single codebook is required to works well for all choices of subsets and message PMFs.

Definition 8 (Semantically-Secure Codes): Let $\alpha \in [0, 1]$ and $\mu = \lfloor \alpha n \rfloor$, a sequence of (n, R) WTC II codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ is α-semantically-secure if there is a constants $\gamma > 0$ and an $n_0 \in \mathbb{N}$, such that for every $n > n_0$, $\mathrm{Sem}_\mu(\mathcal{C}_n) \leq e^{-n\gamma}$.

Definition 9 (SS-Achievability): Let $\alpha \in [0, 1]$ and $\mu = \lfloor \alpha n \rfloor$, a rate $R \in \mathbb{R}_+$ is α-SS-achievable if there is a sequence of (n, R) α-semantically-secure WTC II codes $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ with $e^\star(\mathcal{C}_n) \to 0$ as $n \to \infty$.

Definition 10 (SS-Capacity): For any $\alpha \in [0, 1]$, the α-SS-capacity of the WTC II $C_{\mathrm{Sem}}(\alpha)$ is the supremum of the set of α-SS-achievable rates.

B. Results

The following proposition is subsequently used for the converse proof of the WTC II SS-capacity. The proposition states that the strong-secrecy-capacity of a WTC I with a DM-EC to the eavesdropper is an upper bound on the strong-secrecy-capacity of the WTC II. To formulate the result, slight modifications of some of the definitions from Sections IV-A and V-A are required. Specifically, we redefine the achievable rates for each setting with respect to a strong-secrecy requirement (instead of SS).

Definition 11 (Strong-Secrecy Achievability for WTC I): A rate $R \in \mathbb{R}_+$ is strong-secrecy-achievable for the WTC I if there is a sequence of (n, R) codes $\{\mathcal{C}_{1,n}\}_{n \in \mathbb{N}}$ with

$$e^\star(\mathcal{C}_{1,n}) \xrightarrow[n \to \infty]{} 0 \qquad (67a)$$

$$I_{\mathcal{C}_{1,n}}(M; \mathbf{Z}) \xrightarrow[n \to \infty]{} 0, \qquad (67b)$$

where M is uniformly distributed over the message set $\mathcal{M}$.

Definition 12 (Strong-Secrecy Achievability for WTC II): Let $\alpha \in [0, 1]$ and $\mu = \lfloor \alpha n \rfloor$, a rate $R \in \mathbb{R}_+$ is α-strong-secrecy-achievable for the WTC II if there is a sequence of (n, R) codes $\{\mathcal{C}_{2,n}\}_{n \in \mathbb{N}}$ with

$$e^\star(\mathcal{C}_{2,n}) \xrightarrow[n \to \infty]{} 0 \qquad (68a)$$

$$\max_{\substack{\mathcal{S} \subseteq [1:n]: \\ |\overline{\mathcal{S}}|=\mu}} I_{\mathcal{C}_{2,n}}(M; \mathbf{Z}) \xrightarrow[n \to \infty]{} 0, \qquad (68b)$$

where M is uniformly distributed over the message set $\mathcal{M}$.

The *strong-secrecy-capacity* for both setting is defined as the supremum of the set of strong-secrecy-achievable rates.

Proposition 1 (WTC I Upper Bounds WTC II): Let $\alpha \in (0, 1]$ *and* $C_S^{II}(\alpha)$ *be the* α*-strong-secrecy-capacity of the WTC II with a main channel* $Q_{Y|X}^{(2)}$. *Furthermore, let* $\beta \in [0, \alpha)$ *and* $C_S^I(\beta)$ *be the strong-secrecy-capacity of the WTC I with transition probability* $Q_{Y,Z|X}^{(1)} = Q_{Y|X}^{(2)} \mathcal{E}_{Z|X}^{(\beta)}$, *where* $\mathcal{E}_{Z|X}^{(\beta)}$ *is a DM-EC with erasure probability* $\bar{\beta} = 1 - \beta$, *i.e.,*

$$\mathcal{E}_{Z|X}^{(\beta)}(z|x) = \begin{cases} \beta, & z = x \\ \bar{\beta}, & z = \, ? \end{cases}, \quad \forall x \in \mathcal{X}. \tag{69}$$

Then

$$C_S^{II}(\alpha) \leq C_S^I(\beta) = \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U; Y) - \beta I(U; X) \right]. \tag{70}$$

See Section V-C.1 for the proof. Proposition 1 is subsequently combined with the following lemma to to establish the converse for the α-SS-capacity of the WTC II.

Lemma 6 (Continuity of WTC I Capacity): As a function of β,

$$C_S^I(\beta) = \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U; Y) - \beta I(U; X) \right] \tag{71}$$

is continues inside $(0, 1)$.

The proof of Lemma 6 is relegated to Appendix E. The SS-capacity of the WTC II with a noisy main channel is stated next.

Theorem 2 (WTC II SS-Capacity): For any $\alpha \in [0, 1]$,

$$C_{Sem}(\alpha) = \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U; Y) - \alpha I(U; X) \right], \tag{72}$$

and one may restrict the cardinality of U *to* $|\mathcal{U}| \leq |\mathcal{X}|$.

The converse and direct parts of Theorem 2 are established in Sections V-C.2 and V-C.3, respectively. As oppose to the SS-capacity of the WTC I (where achievability may be derived without using Lemma 1 - see Remark 6), for the WTC II, the stronger soft-covering lemma is essential for the direct proof. Specifically, via the union bound, the double-exponential decay that Lemma 1 provides is leveraged to show the existence of a sequence of codes that satisfies the vanishing information leakage requirement for all choices of $\mathcal{S}$ and P_M.

Remark 9 (Generalized WTC II SS-Capacity): The proof of Theorem 2 is robust and readily extends to a more general setting where the eavesdropper's observed symbols are corrupted by random noise. Specifically, we refer to the scenario where the eavesdropper first chooses a subset of indices $\mathcal{S} \subseteq [1:n]$ *of size* $\mu = \lfloor \alpha n \rfloor$, *then* $\mathbf{x}^{\mathcal{S}}$ *is passed through a DMC* $Q_{Z|X}$ *and the eavesdropper receives* $Z_i \sim Q_{Z|X=x_i}$, *for* $i \in \mathcal{S}$, *and* $Z =?$ *otherwise. The* α*-SS-capacity for this case is*

$$C_{Sem}^{(Noisy)}(\alpha) = \max_{\substack{Q_{U,X}: \\ U-X-(Y,Z)}} \left[I(U; Y) - \alpha I(U; Z) \right], \tag{73}$$

and recovers (72) *by setting* $Z = X$. *Both the direct and the converse proofs of* (73) *follow by a verbatim repetition of the arguments from Section V-C, with two minor changes.*

First, for the converse, the classic DM-EC from Proposition 1 (proven in V-C.1) is replaced with a cascade of the DM-EC and the DMC $Q_{Z|X}$. Second, for the SS analysis in the direct proof (Section V-C.3) we replace the rate bound from (110) with $\tilde{R} > \alpha I(U; Z)$ (the reliability analysis goes through without changes).

Remark 10: The cardinality bound in Theorem 2 is established using the convex cover method [35, Appendix C]. The details are omitted.

Remark 11: Theorem 2 recovers the achievability result from [8, eq. (7)] by setting $U = X$ *and taking X to be uniformly distributed over* $\mathcal{X}$. *Furthermore, in [8] secrecy was established while assuming a uniform distribution over the message set, i.e., on average over the messages. Although we require security with respect to a stricter metric (SS versus weak-secrecy), we achieve higher rates than [8, eq. (7)] and show their optimality. Moreover, to achieve* (72), *we use classic wiretap codes and establish SS using the stronger soft-covering lemma, making the (rather convoluted) coset coding scheme from [8] (inspired by [7]) no longer required.*

C. Proofs

1) Proposition 1: The equality in (70) follows by evaluating the strong-secrecy-capacity formula of a general WTC I, i.e.,

$$\max_{\substack{Q_{U,X}: \\ U-X-(Y,Z)}} \left[I(U; Y) - I(U; Z) \right], \tag{74}$$

for the case where the transition probability matrix is $Q_{Y,Z|X}^{(1)} = Q_{Y|X}^{(2)} \mathcal{E}_{Z|X}^{(\beta)}$. Let $\Phi \sim \text{Ber}(\bar{\beta})$ be a random variable, such that its i.i.d. samples define the erasure process of the DM-EC with erasure probability $\bar{\beta}$. Accordingly, Φ is independent of X and

$$Z = \begin{cases} X, & \Phi = 0 \\ ?, & \Phi = 1. \end{cases} \tag{75}$$

First note that Φ is determined by Z since $? \notin \mathcal{X}$. Combining this with the Markov relation $U - X - (Y, Z)$ implies that the chain $U - X - (Y, Z, \Phi)$ is also Markov. Along with the independence of X and Φ, this implies that U and Φ are also independent. Consequently, for every $Q_{U,X}$, where $U - X - (Y, Z)$ forms a Markov chain, we have

$$\begin{aligned} I(U; Z) &\stackrel{(a)}{=} I(U; \Phi, Z) \\ &\stackrel{(b)}{=} I(U; Z|\Phi) \\ &\stackrel{(c)}{=} \beta I(U; X) + \bar{\beta} I(U; ?) \\ &= \beta I(U; X), \end{aligned} \tag{76}$$

where (a) follows since Φ is defined by Z, while (b) and (c) follows by the independence of Φ and U. Since (76) holds for every $Q_{U,X}$ as above, we conclude that

$$C_S^I(\beta) = \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U; Y) - \beta I(U; X) \right]. \tag{77}$$

To prove the inequality in (70), we show that for any $\alpha \in (0, 1]$ and $\beta \in [0, \alpha)$, an α-strong-secrecy-achievable rate for

the WTC II is also achievable for the WTC I with erasure probability $\bar{\beta}$.

Fix α, β as above and let $R \in \mathbb{R}_+$ be an α-strong-secrecy-achievable rate for the WTC II. Furthermore, let $\{\mathcal{C}_{2,n}\}_{n\in\mathbb{N}}$ be the corresponding sequence of (n, R) codes satisfying (68). Since the channel to the legitimate receiver and the definition of the maximal error probability are the same for both versions of the WTC (see (67a) and (68a)), $\{\mathcal{C}_{2,n}\}_{n\in\mathbb{N}}$ is also reliable when using it to transmit over the WTC I. Therefore, to establish (70), it suffices to show that for every $\epsilon > 0$, there is an $n^\star \in \mathbb{N}$, such that for every $n > n^\star$

$$I_{\mathcal{C}_{2,n}}(M; \mathbf{Z}_1) \leq \epsilon, \qquad (78)$$

where $\mathbf{Z}_1$ denoted the channel output sequences observed by the eavesdroppers of the WTC I. In other words, we show that the sequence of codes $\{\mathcal{C}_{2,n}\}_{n\in\mathbb{N}}$, designed to achieve strong-secrecy for the WTC II, also achieves strong-secrecy for the WTC I.

Let $\mathbf{Z}_2$ be the channel output observed by the eavesdroppers of the WTC II, fix $\epsilon > 0$ and let $n_0 \in \mathbb{N}$ be such that for every $n > n_0$,

$$\max_{\substack{\mathcal{S}\subseteq[1:n]: \\ |\mathcal{S}|=\mu}} I_{\mathcal{C}_{2,n}}(M; \mathbf{Z}_2) \leq \frac{\epsilon}{2}. \qquad (79)$$

For every $\mathbf{z} \in \mathcal{Z}^n$, where $\mathcal{Z} \triangleq \mathcal{X} \cup \{?\}$, define

$$\mathcal{A}(\mathbf{z}) \triangleq \{i \in [1:n] | z_i =?\}, \qquad (80)$$

and let $\Theta(\mathbf{Z})$ be

$$\Theta(\mathbf{Z}) \triangleq \mathbb{1}_{\left\{|\mathcal{A}(\mathbf{Z})| \leq \lceil \bar{\alpha} n \rceil\right\}}. \qquad (81)$$

Namely, Θ indicates if the number of erasures in a sequence $\mathbf{z} \in \mathcal{Z}^n$ is greater than or equal to $\lceil \bar{\alpha} n \rceil$ or not.

By conditioning the mutual information term from (78) on $\Theta(\mathbf{Z}_1)$, we distinguish between the two cases of $\mathbf{Z}_1$ being better or worse than $\mathbf{Z}_2$ in terms of the number of erased symbols. When $\Theta(\mathbf{Z}_1) = 0$, i.e., $\mathbf{Z}_1$ is worse that $\mathbf{Z}_2$, security for the WTC I is ensured since $\{\mathcal{C}_{2,n}\}_{n\in\mathbb{N}}$ achieve security for the WTC II. Otherwise, for the case that $\Theta(\mathbf{Z}_1) = 1$, where $\mathbf{Z}_1$ is better than $\mathbf{Z}_2$, we use Sanov's Theorem to show that the probability of such an event exponentially decreases with the blocklength n, while the mutual information grows linearly at most. For any $n \in \mathbb{N}$, we have

$$
\begin{aligned}
I_{\mathcal{C}_{2,n}}(M; \mathbf{Z}_1) &\overset{(a)}{=} I_{\mathcal{C}_{2,n}}\left(M; \Theta(\mathbf{Z}_1), \mathbf{Z}_1\right) \\
&\overset{(b)}{=} I_{\mathcal{C}_{2,n}}\left(M; \mathbf{Z}_1 \big| \Theta(\mathbf{Z}_1)\right) \\
&= \underbrace{\mathbb{P}\left(\Theta(\mathbf{Z}_1) = 0\right) I_{\mathcal{C}_{2,n}}\left(M; \mathbf{Z}_1 \big| \Theta(\mathbf{Z}_1) = 0\right)}_{\mathcal{I}_0} \\
&\quad + \underbrace{\mathbb{P}\left(\Theta(\mathbf{Z}_1) = 1\right) I_{\mathcal{C}_{2,n}}\left(M; \mathbf{Z}_1 \big| \Theta(\mathbf{Z}_1) = 1\right)}_{\mathcal{I}_1},
\end{aligned}
$$
$$(82)$$

where (a) is because $\Theta(\mathbf{Z}_1)$ is a function of $\mathbf{Z}_1$, while (b) follows since the number of erasures in the output sequence

of a DM-EC is defined by an i.i.d. process that is independent of the message.

For $\mathcal{I}_0$, taking any $n > n_0$, (79) implies that

$$I_{\mathcal{C}_{2,n}}\left(M; \mathbf{Z}_1 \big| \Theta(\mathbf{Z}_1) = 0\right) \leq \max_{\substack{\mathcal{S}\subseteq[1:n]: \\ |\mathcal{S}|=\mu}} I_{\mathcal{C}_{2,n}}(M; \mathbf{Z}_2) \leq \frac{\epsilon}{2}. \quad (83)$$

To upper bound $\mathcal{I}_1$, first note that

$$I_{\mathcal{C}_{2,n}}\left(M; \mathbf{Z}_1 \big| \Theta(\mathbf{Z}_1) = 1\right) \leq n \log\left(|\mathcal{X}| + 1\right), \qquad (84)$$

holds for every $n \in \mathbb{N}$. Now, fix any $\delta \in (\beta, \alpha)$; there exists an $n_1(\delta) \in \mathbb{N}$, such that for all $n > n_1(\delta)$

$$\lceil \bar{\alpha} n \rceil \leq \bar{\delta} n < \bar{\beta} n. \qquad (85)$$

Thus, for every $n > n_1(\delta)$ Sanov's Theorem [22, Th. 11.4.1] implies

$$\mathbb{P}\left(\Theta(\mathbf{Z}_1) = 1\right) \leq \mathbb{P}\left(|\mathcal{A}(\mathbf{Z}_1)| \leq \bar{\delta} n\right) \leq (n+1)^2 \cdot 2^{-nD_b(\delta, \beta)}, \qquad (86)$$

where $D_b(\delta, \beta) = \alpha \log(\delta/\beta) + \bar{\delta} \log(\bar{\delta}/\bar{\beta})$ is the relative entropy between the PMFs of two binary random variables distributed according to $\mathrm{Ber}(\delta)$ and $\mathrm{Ber}(\beta)$, respectively. Since $\delta \neq \beta$, we have that $D_b(\delta, \beta) > 0$, and therefore, there is an $n_1(\delta) < n_2 \in \mathbb{N}$, such that for every $n > n_2$,

$$\mathcal{I}_2 \leq (n+1)^2 \cdot 2^{-nD_b(\delta, \beta)} \cdot n \log\left(|\mathcal{X}| + 1\right) \leq \frac{\epsilon}{2}. \quad (87)$$

Set $n^\star = \max\{n_0, n_2\}$. Based on (83) and (87), for every $n > n^\star$, we have

$$I_{\mathcal{C}_{2,n}}(M; \mathbf{Z}_1) = \mathcal{I}_0 + \mathcal{I}_1 \leq \epsilon, \qquad (88)$$

which completes the proof.

2) Theorem 2 - Converse: For the converse, we first show that with respect to the notations used in Proposition 1,

$$C_S^{\mathrm{II}}(\alpha) \leq C_S^{\mathrm{I}}(\alpha) = \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U; Y) - \alpha I(U; X)\right], \qquad (89)$$

for any $\alpha \in [0, 1]$. For $\alpha = 0, 1$, the relation is straightforward as

$$C_S^{\mathrm{I}}(0) = \max_{Q_X} I(X; Y) = C_S^{\mathrm{II}}(0) \qquad (90\mathrm{a})$$

$$C_S^{\mathrm{I}}(1) = 0 = C_S^{\mathrm{II}}(1). \qquad (90\mathrm{b})$$

For $\alpha \in (0, 1)$, (89) is established by relying on Proposition 1 and the continuity argument from Lemma 6. Namely, by taking the limit of (70) as $\beta \uparrow \alpha$ establishes (89).

Having this, the converse follows by arguments similar to those presented in Section IV-C.1. Fix $\alpha \in [0, 1]$ and let $R \in \mathbb{R}_+$ be an α-SS-achievable rate for the WTC II and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be its corresponding (n, R) sequence of codes. By the definitions in (42a) and (66), $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ are reliable and α-semantically-secure for every message distribution, and in particular, for a uniform message distribution. This implies

$$C_{\mathrm{Sem}}(\alpha) \leq C_S^{\mathrm{II}}(\alpha) \leq \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U; Y) - \alpha I(U; X)\right] \qquad (91)$$

and completes the converse proof.

Remark 12: Our converse proof relies on the achievability being defined in terms of a limit as $n \rightarrow \infty$ (see Definition 9). Namely, we show that in the limit, the eavesdropper in the WTC I setting is likely to be within a slightly higher channel-observation budget than this of the WTC II, which by continuity won't result in much extra rate. The chance of having too many channel observations is too small to provide non-negligible extra information. If, however, the blocklength n can be chosen as a design parameter, then it may be possible that a finite n results in a higher achievable secrecy-rate. For instance, notice that the optimal code of length $2n$ in not necessarily better than the optimal code of length n, since when the blocklenght is longer the eavesdropper has more flexibility in choosing his observations.

3) Theorem 2 - Direct Part: As before, we start by showing the achievability of (72) when $U = X$. After doing so, we use channel prefixing to extend the proof to any U with $U-X-Y$.

Fix $\alpha \in [0,1]$, $\epsilon > 0$ and a PMF Q_X on $\mathcal{X}$. Letting M and W be independent random variables uniformly distributed over $\mathcal{M}$ and $\mathcal{W} = [1:2^{n\tilde{R}}]$, respectively, we repeat the code construction from Section IV-C.1. A similar analysis of the average error probability shows that if

$$R + \tilde{R} < I(X;Y), \tag{92}$$

then

$$\mathbb{E}_{\mathbb{B}_n} \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} e_m(\mathbb{C}_n) \xrightarrow[n\to\infty]{} 0, \tag{93}$$

where $\mathbb{C}_n$ is the random code that corresponds to the random codebook $\mathbb{B}_n$.

Security Analysis: Fix $\mathcal{S} \subseteq [1:n]$ with $|\mathcal{S}| = \mu = \lfloor \alpha n \rfloor$, recall that $\mathcal{Z} \triangleq \mathcal{X} \cup \{?\}$ and define the following PMF on $\mathcal{Z}^n$,

$$\Gamma_{\mathbf{Z}}^{(\mathcal{S})}(\mathbf{z}) = \prod_{j \in \mathcal{S}^c} \mathbb{1}_{\{z_j = ?\}} \prod_{j \in \mathcal{S}} \mathcal{I}_Z(z_j), \quad \forall \mathbf{z} \in \mathcal{Z}^n, \tag{94}$$

where $\mathcal{I}_Z$ is the average output PMF of the identity DMC on $\mathcal{X}$, i.e.,

$$\mathcal{I}_Z(z) = \sum_{x \in \mathcal{X}} Q_X(x) \mathbb{1}_{\{z=x\}} = \begin{cases} Q_X(z), & z \in \mathcal{X} \\ 0, & z =?. \end{cases} \tag{95}$$

For any $\mathcal{C}_n$ (defined by fixing $\mathcal{B}_n$) and $P_M \in \mathcal{P}(\mathcal{M})$, the relative entropy chain rule implies

$$I_{\mathcal{C}_n}(M;\mathbf{Z}) = D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n,\mathcal{S})} \middle\| P_{\mathbf{Z}}^{(\mathcal{C}_n,\mathcal{S})} \middle| P_M\right)$$
$$= D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}}^{(\mathcal{S})} \middle| P_M\right)$$
$$- D\left(P_{\mathbf{Z}}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}}^{(\mathcal{S})} \middle| P_M\right), \tag{96}$$

and therefore

$$\max_{P_M \in \mathcal{P}(\mathcal{M})} I_{\mathcal{C}_n}(M;\mathbf{Z})$$
$$\leq \max_{P_M \in \mathcal{P}(\mathcal{M})} D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}}^{(\mathcal{S})} \middle| P_M\right)$$
$$\leq \max_{P_M \in \mathcal{P}(\mathcal{M})} \sum_{m \in \mathcal{M}} P(m) \max_{\tilde{m} \in \mathcal{M}} D\left(P_{\mathbf{Z}|M=\tilde{m}}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}}^{(\mathcal{S})}\right)$$
$$= \max_{m \in \mathcal{M}} D\left(P_{\mathbf{Z}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}}^{(\mathcal{S})}\right). \tag{97}$$

For any $\emptyset \neq \mathcal{A} \subseteq [1:n]$ and $\mathbf{z} \in \mathcal{Z}^n$, recall that $\mathbf{z}^{\mathcal{A}} \triangleq (z_i)_{i \in \mathcal{A}}$ is the sub-vector of $\mathbf{z}$ indexed by the elements of $\mathcal{A}$. The relative entropy chain rule further simplifies the RHS of (97) as follows. For any $m \in \mathcal{M}$, we have

$$D\left(P_{\mathbf{Z}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}}^{(\mathcal{S})}\right) = D\left(P_{\mathbf{Z}^{\mathcal{S}},\mathbf{Z}^{\mathcal{S}^c}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}^{\mathcal{S}},\mathbf{Z}^{\mathcal{S}^c}}^{(\mathcal{S})}\right)$$
$$= D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}^{\mathcal{S}}}^{(\mathcal{S})}\right)$$
$$+ D\left(P_{\mathbf{Z}^{\mathcal{S}^c}|M=m,\mathbf{Z}^{\mathcal{S}}}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}^{\mathcal{S}^c}}^{(\mathcal{S})} \middle| P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathcal{C}_n,\mathcal{S})}\right)$$
$$\stackrel{(a)}{=} D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \Gamma_{\mathbf{Z}^{\mathcal{S}}}^{(\mathcal{S})}\right)$$
$$\stackrel{(b)}{=} D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^{\mu}\right), \tag{98}$$

where (a) is because $P_{\mathbf{Z}^{\mathcal{S}^c}|M=m,\mathbf{Z}^{\mathcal{S}}=\mathbf{z}^{\mathcal{S}}}^{(\mathcal{C}_n,\mathcal{S})} = \mathbb{1}_{\{Z_i=?, \ i \in \mathcal{S}^c\}} = \Gamma_{\mathbf{Z}^{\mathcal{S}^c}}^{(\mathcal{S})}$, for every $\mathbf{z}^{\mathcal{S}} \in \mathcal{Z}^{|\mathcal{S}|}$, and (b) follows from (94).

Combining (96)-(98), we have that for every $\mathcal{C}_n$ and $\mathcal{S} \subseteq [1:n]$, with $|\mathcal{S}| = \mu = \lfloor \alpha n \rfloor$,

$$\max_{P_M \in \mathcal{P}(\mathcal{M})} D\left(P_{\mathbf{Z}|M}^{(\mathcal{C}_n,\mathcal{S})} \middle\| P_{\mathbf{Z}}^{(\mathcal{C}_n,\mathcal{S})} \middle| P_M\right)$$
$$\leq \max_{m \in \mathcal{M}} D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^{\mu}\right). \tag{99}$$

In particular, (99) also holds when maximizing over the substes S, which gives

$$\text{Sem}_\mu(\mathcal{C}_n) \leq \max_{\substack{m \in \mathcal{M}, \\ \mathcal{S} \subseteq [1:n]: \ |\mathcal{S}|=\mu}} D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^{\mu}\right). \tag{100}$$

Having (100), let $\tilde{\delta}$ be an arbitrary positive real number to be determined later and consider the following probability.

$$\mathbb{P}\left(\left\{\text{Sem}_\mu(\mathbb{C}_n) \leq e^{-n\tilde{\delta}}\right\}^c\right)$$
$$= \mathbb{P}\left(\max_{\substack{P_M \in \mathcal{P}(\mathcal{M}), \\ \mathcal{S} \subseteq [1:n]: \ |\mathcal{S}|=\mu}} D\left(P_{\mathbf{Z}|M}^{(\mathbb{C}_n,\mathcal{S})} \middle\| P_{\mathbf{Z}}^{(\mathbb{C}_n,\mathcal{S})} \middle| P_M\right) > e^{-n\tilde{\delta}}\right)$$
$$\stackrel{(a)}{\leq} \mathbb{P}\left(\max_{\substack{m \in \mathcal{M}, \\ \mathcal{S} \subseteq [1:n]: \ |\mathcal{S}|=\mu}} D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathbb{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^{\mu}\right) > e^{-n\tilde{\delta}}\right)$$
$$= \mathbb{P}\left(\bigcup_{\substack{m \in \mathcal{M}, \\ \mathcal{S} \subseteq [1:n]: \ |\mathcal{S}|=\mu}} \left\{D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathbb{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^{\mu}\right) > e^{-n\tilde{\delta}}\right\}\right)$$
$$\stackrel{(b)}{\leq} \sum_{\substack{m \in \mathcal{M}, \\ \mathcal{S} \subseteq [1:n]: \ |\mathcal{S}|=\mu}} \mathbb{P}\left(D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathbb{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^{\mu}\right) > e^{-n\tilde{\delta}}\right), \tag{101}$$

where (a) uses (100), and (b) is the union bound.

Each term in the sum on the RHS of (101) falls into the framework of the stronger soft-covering lemma, with respect to a blocklength of μ and the identity channel. Noting that $|\mathcal{W}| = 2^{n\tilde{R}} = 2^{\mu \frac{n\tilde{R}}{\mu}}$, we have that as long as

$$\frac{n\tilde{R}}{\mu} > H(X), \tag{102}$$

there exist $\delta_1, \delta_2 > 0$ that for sufficiently large n satisfy

$$\mathbb{P}\left(D\left(P_{\mathbf{Z}^{\mathcal{S}}|M=m}^{(\mathbb{C}_n,\mathcal{S})} \middle\| \mathcal{I}_Z^\mu \right) > e^{-n\delta_1} \right) \le e^{-e^{n\delta_2}}. \qquad (103)$$

Since $\mu = \lfloor \alpha n \rfloor \le \alpha n$, taking

$$\tilde{R} > \alpha H(X), \qquad (104)$$

is sufficient to satisfy (102) for every $n \in \mathbb{N}$.

Setting $\tilde{\delta} = \delta_1$ and plugging (103) into (101), gives

$$\mathbb{P}\left(\left\{ \mathrm{Sem}_\mu(\mathbb{C}_n) \le e^{-n\delta_1} \right\}^c \right) \le \sum_{\substack{m \in \mathcal{M}, \\ \mathcal{S} \subseteq [1:n]:\ |\mathcal{S}|=\mu}} e^{-e^{n\delta_2}}$$

$$\le 2^n \cdot 2^{nR} \cdot e^{-e^{n\delta_2}}$$

$$\triangleq \kappa_n \xrightarrow[n\to\infty]{} 0.$$

Invoking Lemma 5 once more, we have that if (92) and (102) are satisfied, then there is a sequence of (n, R) α-semantically-secure codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, with

$$\mathbb{E}_{\mathbb{C}_n} \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} e_m(\mathbb{C}_n) \xrightarrow[n\to\infty]{} 0. \qquad (105)$$

The pruning argument from Section IV-C.1 again upgrades $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ to be reliable with respect to the maximal error probability. Combining (92) and (102) shows the achievability of

$$R < \max_{Q_X} \left[I(X;Y) - \alpha H(X) \right]. \qquad (106)$$

Finally, we prefix a DMC $Q_{X|U}$ to the original WTC II to obtain a new main channel $Q_{Y|U}$, given by

$$Q_{Y|U}^n(\mathbf{y}|\mathbf{u}) = \sum_{\mathbf{x}\in\mathcal{X}^n} Q_{X|U}^n(\mathbf{x}|\mathbf{u}) Q_{Y|X}^n(\mathbf{y}|\mathbf{x}). \qquad (107)$$

Furthermore, $\Gamma_{\mathbf{Z}}^{(\mathcal{S})}$ from (94) is redefined as

$$\Gamma_{\mathbf{Z}}^{(\mathcal{S})}(\mathbf{z}) = \prod_{j\in\mathcal{S}^c} \mathbb{1}_{\left\{ z_j=? \right\}} \prod_{j\in\mathcal{S}} Q_Z(z_j), \quad \forall \mathbf{z}\in\mathcal{Z}^n, \qquad (108)$$

where Q_Z is given by

$$Q_Z(z) = \sum_{(u,x)\in\mathcal{U}\times\mathcal{X}} Q_U(u) Q_{X|U}(x|u) \mathbb{1}_{\{z=x\}}$$

$$= \begin{cases} \sum_{u\in\mathcal{V}} Q_U(u) Q_{X|U}(z|u), & z\in\mathcal{X} \\ 0, & z=?. \end{cases}$$

Repeating a similar analysis as above shows that reliability is achieved if

$$R + \tilde{R} < I(U;Y), \qquad (109)$$

while the rate needed for the stronger soft-covering lemma is

$$\tilde{R} > \alpha I(U;X). \qquad (110)$$

Putting (109)-(110) together yields that any rate $R \in \mathbb{R}_+$ satisfying

$$R < \max_{\substack{Q_{U,X}: \\ U-X-Y}} \left[I(U;Y) - \alpha I(U;X) \right], \qquad (111)$$

is strongly α-SS-achievable and concludes the proof.

VI. Summary and Concluding Remarks

We derived the SS capacity of the WTC II with a noisy main channel. The SS metric ensures that the unnormalized mutual information between the message and the eavesdropper's observation is arbitrarily small, even when maximized over all message distributions and all possible choices of the eavesdropper's observation. The main tool used in the direct proof is a novel and stronger version of Wyner's soft covering lemma, that states that a random codebook achieves the soft-covering phenomenon with high probability as long as its rate is higher than the mutual information between the input and output of the DMC. Furthermore, the probability of failure is doubly-exponentially small in the blocklength, thus making the lemma advantageous in proving the existence of codebooks that satisfy exponentially many constraints. A code that achieves SS for the considered WTC II should do just that.

The SS capacity was achieved by using classic Wyner's wiretap codes. Since the combined number of messages and subsets grows only exponentially with the blocklength, SS was established by applying the union bound and invoking the stronger soft-covering lemma. The direct proof showed that rates up to the weak-secrecy capacity of the WTC I with a DM-EC to the eavesdropper are achievable. The converse followed by showing that the capacity of this WTC I is an upper bound on the SS capacity of the WTC II.

As a preliminary and simple application of the stronger soft-covering lemma, it was used to achieve SS for the WTC I. A main goal in doing so was to emphasize the advantage of this approach over other methods for achieving SS for this scenario, such as the expurgation technique. While the expurgation method fails to generalize to some multiuser settings, such as the multiple access WTC, an achievability proof that relies on the stronger soft-covering lemma goes through by similar steps to those presented here. Thus making the stronger soft-covering lemma a tool by which the common weak-secrecy and strong-secrecy results can be upgraded to SS. Furthermore, the lemma might prove useful in any other scenario in which performance is measures with respect to an exponential number of constraints.

Appendix A
Proof of Lemma 2

Let $n_0 \in \mathbb{N}$ be such that (6) holds for any $n > n_0$. For these values of n we have

$$\mathbb{E}_{\mathbb{B}_n} D\left(P_{\mathbf{V}}^{(\mathcal{B}_n)} \middle\| Q_V^n \right)$$

$$= \mathbb{E}_{\mathbb{B}_n} \left[D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \middle\| Q_V^n \right) \left(\mathbb{1}_{\left\{ D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \| Q_V^n \right) \le e^{-n\gamma_1} \right\}} \right. \right.$$

$$\left. \left. + \mathbb{1}_{\left\{ D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \| Q_V^n \right) > e^{-n\gamma_1} \right\}} \right) \right]$$

$$\overset{(a)}{\le} e^{-n\gamma_1} + n\log\left(\frac{1}{\mu_V} \right) \mathbb{P}\left(D\left(P_{\mathbf{V}}^{(\mathbb{B}_n)} \middle\| Q_V^n \right) > e^{-n\gamma_1} \right)$$

$$\overset{(b)}{\le} e^{-n\gamma_1} + n\log\left(\frac{1}{\mu_V} \right) e^{-e^{n\gamma_2}}, \qquad (112)$$

where (a) follows because for every fixed $\mathcal{B}_n$

$$D\left(P_{\mathbf{V}}^{(\mathcal{B}_n)}\middle\|Q_V^n\right) = \sum_{\mathbf{z}\in\mathcal{Z}^n} P^{(\mathcal{B}_n)}(\mathbf{z})\log\left(\frac{P^{(\mathcal{B}_n)}(\mathbf{v})}{Q_V^n(\mathbf{v})}\right)$$

$$\leq n\log\left(\frac{1}{\mu_V}\right),$$

and $\mu_V = \min_{v\in\text{supp}(Q_V)} Q_V(v)$, while (b) follows from (6).

APPENDIX B
PROOF OF LEMMA 3

Fix a codebook $\mathcal{B}_n$ and define

$$\Theta = \mathbb{1}_{\left\{\left(\mathbf{U}(W,\mathcal{B}_n),\mathbf{V}\right)\notin\mathcal{A}_\epsilon\right\}} + 1. \tag{113}$$

Note that for $\theta = 1, 2$, we have

$$P_{\Theta}^{(\mathcal{B}_n)}(\theta) = \int dP_{\mathbf{V}}^{(\mathcal{B}_n)} \sum_{w\in\mathcal{W}} 2^{-nR} Q_{V|U=\mathbf{u}(w,\mathcal{B}_n)}^n$$

$$\times\Bigg[\mathbb{1}_{\{\theta=1\}\cap\left\{\left(\mathbf{u}(w,\mathcal{B}_n),\mathbf{V}\right)\in\mathcal{A}_\epsilon\right\}}$$

$$+\mathbb{1}_{\{\theta=2\}\cap\left\{\left(\mathbf{u}(w,\mathcal{B}_n),\mathbf{V}\right)\notin\mathcal{A}_\epsilon\right\}}\Bigg]$$

$$= \int dP_{\mathcal{B}_n,\theta}, \tag{114}$$

and consequently, for every measurable $\mathcal{A}\subseteq\mathcal{V}^n$,

$$\mathbb{P}_{P_{\mathbf{V},\Theta}^{(\mathcal{B}_n)}}\left(\mathbf{V}\in\mathcal{A}, \Theta=\theta\right) = \mathbb{P}_{P_{\mathcal{B}_n,\theta}}\left(\mathbf{V}\in\mathcal{A}\right)$$

$$= \int_{\mathcal{A}} dP_{\mathcal{B}_n,\theta}. \tag{115}$$

For simplicity of notation, denote $P_{\mathbf{V}}^{(\mathcal{B}_n)} \triangleq P$, $P_{\mathcal{B}_n,1} \triangleq P_1$, $P_{\mathcal{B}_n,2} \triangleq P_2$, $Q_V^n \triangleq Q$ and $P_{\Theta}^{(\mathcal{B}_n)} \triangleq \Gamma_\Theta$, and consider

$$D(P\|Q) = \int dP\log\left(\frac{dP}{dQ}\right)$$

$$\stackrel{(a)}{=} \int dQ\frac{dP}{dQ}\log\left(\frac{dP}{dQ}\right)$$

$$\stackrel{(b)}{=} \int dQ\mathbb{E}_{\Gamma_\Theta}\left[\frac{1}{\Gamma_\Theta(\Theta)}\cdot\frac{dP_\Theta}{dQ}\right]$$

$$\times\log\left(\mathbb{E}_{\Gamma_\Theta}\left[\frac{1}{\Gamma_\Theta(\Theta)}\cdot\frac{dP_\Theta}{dQ}\right]\right)$$

$$\stackrel{(c)}{\leq} \int dQ\mathbb{E}_{\Gamma_\Theta}\left[\frac{1}{\Gamma_\Theta(\Theta)}\cdot\frac{dP_\Theta}{dQ}\cdot\right.$$

$$\left.\times\log\left(\frac{1}{\Gamma_\Theta(\Theta)}\cdot\frac{dP_\Theta}{dQ}\right)\right]$$

$$= \sum_{\theta=1,2}\Gamma_\Theta(\theta)\int dQ\frac{1}{\Gamma_\Theta(\theta)}\cdot\frac{dP_\theta}{dQ}$$

$$\times\log\left(\frac{1}{\Gamma_\Theta(\theta)}\cdot\frac{dP_\theta}{dQ}\right)$$

$$\stackrel{(d)}{=} \sum_{\theta=1,2}\log\left(\frac{1}{\Gamma_\Theta(\theta)}\right)\int dP_\theta$$

$$+ \sum_{\theta=1,2}\int dP_\theta\log\left(\frac{dP_\theta}{dQ}\right)$$

$$\stackrel{(e)}{=} h\left(\int dP_1\right) + \sum_{\theta=1,2}\int dP_\theta\log\Delta_{\mathcal{B}_n,\theta}, \tag{116}$$

where:
(a) follows since for any two measures μ, λ with $\mu \ll \lambda$ and a μ−integrable function g, we have

$$\int gd\mu = \int g\frac{d\mu}{d\lambda}d\lambda; \tag{117}$$

(b) follows from (115) and the law of total probability;
(c) follows by applying Jensen's inequality to the convex function $x \mapsto x\log(x)$;
(d) follows by the properties of the logarithm and (117);
(e) follows from (114) and the definition of $\Delta_{\mathcal{B}_n,\theta}$, for $\theta = 1, 2$, in (17).

APPENDIX C
PROOF OF THE CHENOFF BOUND - LEMMA 4

Let X have the same distribution as X_1. For any $\lambda > 0$, we have

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M}X_m \geq c\right) \stackrel{(a)}{\leq} \frac{\mathbb{E}e^{\lambda\sum_{m=1}^{M}X_m}}{e^{\lambda cM}}$$

$$= \left(\frac{\mathbb{E}e^{\lambda X}}{e^{\lambda c}}\right)^M$$

$$\stackrel{(b)}{\leq} \left(\frac{1+\frac{e^{\lambda B}-1}{B}\mathbb{E}X}{e^{\lambda c}}\right)^M$$

$$\leq \left(\frac{1+\frac{e^{\lambda B}-1}{B}\mu}{e^{\lambda c}}\right)^M$$

$$\stackrel{(c)}{\leq} \left(\frac{e^{\frac{e^{\lambda B}-1}{B}\mu}}{e^{\lambda c}}\right)^M$$

$$= e^{-M\left(\lambda c+\frac{\mu}{B}-\frac{\mu e^{\lambda B}}{B}\right)}, \tag{118}$$

where (a) is the Chernoff bound, (b) uses the fact that $e^{\lambda x} \leq 1 + \frac{e^{\lambda B}-1}{B}x$, for $x \in [0, B]$ due to convexity, and (c) uses $1 + x \leq e^x$.

Optimizing the RHS of (118) over λ given $\lambda^\star = \frac{1}{B}\ln\frac{c}{\mu}$ as the minimizer, as long as $\frac{c}{\mu} \geq 1$. Plugging this into (118) yields

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M}X_m \geq c\right) \leq e^{-\frac{M\mu}{B}\left(\frac{c}{\mu}\left(\ln\frac{c}{\mu}-1\right)+1\right)}, \quad \forall\frac{c}{\mu} \geq 1. \tag{119}$$

This is a good bound when $\mu \ll B$, as it is in our case. If c/μ is shrinking, then to further simplify the bound consider the third order Taylor expansion of $x(\ln x - 1)$ about $x = 1$,

$$x(\ln x - 1) + 1 \geq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3, \quad \forall x \geq 1. \tag{120}$$

The LHS in (120) is a lower bound because the fourth derivative is positive for all $x \geq 1$. Furthermore, if $x - 1 \leq 1$, we have

$$\frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3 \geq \frac{1}{3}(x-1)^2, \quad \forall x \in [1, 2]. \tag{121}$$

Putting it all together gives

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M} X_m \geq c\right) \leq e^{-\frac{M\mu}{3B}\left(\frac{c}{\mu}-1\right)^2}, \quad \forall \frac{c}{\mu} \in [1, 2]. \quad (122)$$

APPENDIX D
PROOF OF LEMMA 5

Since $\left\{f_n^{(i)}\right\}_{n\in\mathbb{N}}$, $i \in [1 : I]$, are bounded and by (57a), there exists a sequence $\{\delta_n\}_{n\in\mathbb{N}}$ such that

$$\mathbb{E}f_n^{(i)}(A_n) \leq \delta_n, \quad \forall i \in [1 : I], \ n \in \mathbb{N}, \quad (123)$$

and $\delta_n \to 0$ as $n \to \infty$. We have

$$\mathbb{P}\left(\bigcup_{i=1}^{I}\left\{f_n^{(i)}(A_n) \geq (I+1)\delta_n\right\}\right)$$

$$\leq \sum_{i=1}^{I}\mathbb{P}\left(f_n^{(i)}(A_n) \geq (I+1)\delta_n\right)$$

$$\overset{(a)}{\leq} \sum_{i=1}^{I}\frac{\mathbb{E}f_n^{(i)}(A_n)}{(I+1)\delta_n}$$

$$\overset{(b)}{\leq} \frac{I}{(I+1)}$$

$$< 1. \quad (124)$$

Therefore, there exists a realization $\{a_n\}_{n\in\mathbb{N}}$ of $\left\{A_n\right\}_{n\in\mathbb{N}}$ such that

$$f_n^{(i)}(a_n) < (I+1)\delta_n \triangleq \tilde{\delta}_n, \quad \forall i \in [1 : I], \quad n \in \mathbb{N}. \quad (125)$$

Since $I < \infty$ independently of n, we have $\tilde{\delta}_n \to 0$ as $n \to \infty$.

APPENDIX E
PROOF OF LEMMA 6

We prove the continuity of $C_S^I(\beta)$ inside $(0, 1)$ by showing that it is bounded and convex. Let $\beta_1, \beta_2 \in (0, 1)$, $\lambda \in [0, 1]$ and observe that

$$C_S^I(\lambda\beta_1 + \bar{\lambda}\beta_2)$$

$$= \max_{\substack{Q_{U,X}:\\U-X-Y}}\left[(\lambda + \bar{\lambda})I(U;Y) - (\lambda\beta_1 + \bar{\lambda}\beta_2)I(U;X)\right]$$

$$\leq \lambda \max_{\substack{Q_{U,X}:\\U-X-Y}}\left[I(U;Y) - \beta_1 I(U;X)\right]$$

$$+ \bar{\lambda} \max_{\substack{Q_{U,X}:\\U-X-Y}}\left[I(U;Y) - \beta_2 I(U;X)\right]$$

$$= \lambda C_S^I(\beta_1) + \bar{\lambda}C_S^I(\beta_2). \quad (126)$$

Furthermore, for every $\beta \in (0, 1)$,

$$C_S^I(\beta) \leq \max_{Q_X} I(X;Y) \leq \log|\mathcal{Y}| < \infty. \quad (127)$$

ACKNOWLEDGMENT

REFERENCES

[1] M. Bellare and S. Tessaro, "Polynomial-time, semantically-secure encryption achieving the secrecy capacity," Cryptol. ePrint Arch., Tech. Rep. 2012/022, 2012. [Online]. Available: http://eprint.iacr.org/

[2] M. Bellare, S. Tessaro, and A. A. Vardy, "A cryptographic treatment of the wiretap channel," in *Proc. Adv. Cryptol. (CRYPTO)*, Santa Barbara, CA, USA, Aug. 2012, pp. 1–31.

[3] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comput. Syst. Sci.*, vol. 28, no. 2, pp. 270–299, Apr. 1984.

[4] M. Bellare, A. Desai, E. Jokipii, and P. A. Rogaway, "A concrete security treatment of symmetric encryption," in *Proc. IEEE 38th Annu. Symp. Found. Comput. Sci.*, Miami, FL, USA, Oct. 1997, pp. 394–403.

[5] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[6] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[7] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," *AT T Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, Dec. 1984.

[8] M. Nafea and A. Yener, "Wiretap channel II with a noisy main channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 1159–1163.

[9] M. J. Mihaljević, *On Message Protection in Cryptosystems Modelled as the Generalized Wire-Tap Channel II* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 1994, pp. 13–24.

[10] Y. Luo, C. Mitrpant, A. J. H. Vinck, and K. Chen, "Some new characters on the wire-tap channel of type II ," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1222–1229, Mar. 2005.

[11] R. Liu, Y. Liang, H. V. Poor, and P. Spasojevic, "Secure nested codes for type II wiretap channels," in *Proc. Inf. Theory Workshop (ITW)*, Lake Tahoe, CA, USA, Sep. 2007, pp. 337–342.

[12] V. Aggarwal, L. Lai, A. R. Calderbank, and H. V. Poor, "Wire-tap channel type II with an active eavesdropper," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun./Jul. 2009, pp. 1944–1948.

[13] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[14] E. C. Song, P. Cuff, and H. V. Poor, "A rate-distortion based secrecy system with side information at the decoders," in *Proc. 52nd Annu. Allerton Conf. Commun., Control Comput.*, Monticell, IL, USA, Sep./Oct. 2014, pp. 755–762.

[15] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

[16] C. Schieler and P. Cuff, "The henchman problem: Measuring secrecy by the minimum distortion in a list," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 596–600.

[17] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7584–7605, Dec. 2014.

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

[19] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[20] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy source compression," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 2042–2046.

[21] M. Hayashi and R. Matsumoto, "Secure multiplex coding with dependent and non-uniform multiple messages," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2355–2409, May 2016.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[23] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.

[24] H. Tyagi and A. Vardy. (2014). "Semantically-secure coding scheme achieving the capacity of a Gaussian wiretap channel." [Online]. Available: http://arxiv.org/abs/1412.4958v2

[25] V. K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. Inf. Theory*, vol. 37, no. 5, pp. 1412–1418, Sep. 1991.

[26] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J.-M. Merolla, "Applications of LDPC codes to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2933–2945, Aug. 2007.

[27] J. L. Massey, "Applied digital information theory," ETH Zürich, Zürich, Switzerland, Tech. Rep., 1998.

[28] 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, pp. 1562–1575, Apr. 2006.

[29] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 569–579, Mar. 2002.

[30] M. M. Wilde, *Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[31] A. Winter, "Secret, public and quantum correlation cost of triples of random variables," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Adelaide, SA, Australia, Sep. 2005, pp. 2270–2274.

[32] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *Proc. 13th Can. Workshop Inf. Theory (CWIT)*, Toronto, ON, Canada, Jun. 2013, pp. 76–81.

[33] P. Cuff, "Soft covering with high probability," *Accepted to IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[34] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3989–4001, Jun. 2011.

[35] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[36] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge, U.K.: Cambridge Univ. Press, Oct. 2011.

Ziv Goldfeld (S'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben- Gurion University, Israel, in 2012 and 2014, respectively. He is currently a student in the direct Ph.D. program for honor students in Electrical and Computer Engineering at that same institution.

Between 2003 and 2006, he served in the intelligence corps of the Israeli Defense Forces.

Ziv is a recipient of several awards, among them the Dean's List Award, the Basor Fellowship for honor students in the direct Ph.D. program, the Lev-Zion fellowship and the Minerva Short-Term Research Grant (MRG).

Paul Cuff (S'08–M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 2006 and 2009. Since 2009 he has been an Assistant Professor of Electrical Engineering at Princeton University.

As a graduate student, Dr. Cuff was awarded the ISIT 2008 Student Paper Award for his work titled Communication Requirements for Generating Correlated Random Variables and was a recipient of the National Defense Science and Engineering Graduate Fellowship and the Numerical Technologies Fellowship. As faculty, he received the NSF Career Award in 2014 and the AFOSR Young Investigator Program Award in 2015.

Haim H. Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial board of the IEEE TRANSACTIONS ON INFORMATION THEORY.