Semantic-Security Capacity for the Physical Layer via Information Theory

Ziv Goldfeld
Ben Gurion University
gziv@post.bgu.ac.il

Paul Cuff
Princeton University
cuff@princeton.edu

Haim H. Permuter
Ben Gurion University
haimp@bgu.ac.il

***Abstract*—Physical layer security can ensure secure communication over noisy channels in the presence of an eavesdropper with unlimited computational power. We adopt an information theoretic variant of semantic-security (SS) (a cryptographic gold standard), as our secrecy metric and study the open problem of the type II wiretap channel (WTC II) with a noisy main channel is, whose secrecy-capacity is unknown even under looser metrics than SS. Herein the secrecy-capacity is derived and shown to be equal to its 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. On top of that, SS requires 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, the lemma shows that a random codebook achieves 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, physical-layer security, semantic-security, soft-covering lemma, wiretap channel of type II, wiretap codes.**

I. Introduction

Modern communication systems usually present an architectural separation between error correction and data encryption. The former is typically realized at the physical-layer by transforming the noisy communication channel into a reliable "bit pipe". The data encryption is implemented on top of that by applying cryptographic principles. The cryptographic approach assumes no knowledge on the quality of the eavesdropper's channel and relies solely on restricting the computational power of the eavesdropper. However, as the construction of quantum computers edges closer (several companies have recently reported a working prototype of a 512-qubit quantum computer), the validity of the restricted computational power assumption comes into question. Nonetheless, cryptography remains the main practical tool for protecting data, at least for the time being.

An alternative approach to secure communication is the so-called physical layer security (or information theoretic security), a concept that dates back to Wyner's celebrated paper on the wiretap channel (WTC) [1]. Essentially, Wyner's main idea was to exploit the noise of the communication channel along with proper physical-layer coding to guarantee secrecy against a computationally-unlimited eavesdropper.

Information theoretic security has adopted the weak- and strong-secrecy metrics as a standard for measuring security. Respectively, weak- 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 [2], [3]. Their main drawback lies is 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) [4], [5] is a cryptographic gold standard originally designed to ensure that a computationally bounded adversary cannot extract secret information in scenarios when information theoretic security is impossible. Recently, an information theoretic adaptation of SS that accounts for computationally unbounded adversaries was proposed in [3] as an adequate alternative to the existing information theoretic security metrics. The authors of [3] showed that this information theoretic SS (henceforth simply referred to as SS) is 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 a noisy main channel, for which even the weak-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 [1]. 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

978-1-5090-1018-9/16 $31.00 © 2016 IEEE
DOI 10.1109/SWSTE.2016.12

the transmitted symbols that 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. In [7], the rate-equivocation region for the case where the main channel is *noiseless* was established by using a unique randomized coset coding scheme in the proof of achievability. 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 in [8] 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 case where equivocation is maximal 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. Wyners original soft-covering lemma [13, Theorem 6.3] is a valuable tool for achievability proofs of information theoretic security [14]–[16], resolvability [17], channel synthesis [18], and source coding [19] (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.

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. The SS-capacity under an average error probability constraint was established in [3] for the DM scenario and in [20] for the Gaussian case. These works presented efficient code constructions with polynomial complexity. 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 insures 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 well-known expurgation technique (e.g., cf. [21, Theorem 7.7.1]). The derivation of the WTC I SS-capacity

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. Using the expurgation method once more establishes achievability and 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 [22], [23].

II. Notations and Preliminaries

We use the following notations. Given two real numbers a, b, we denote by $[a:b]$ the set of integers $\big\{n \in \mathbb{N} \big| \lceil a \rceil \leq n \leq \lfloor b \rfloor \big\}$. 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 $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ 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 $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ 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 $\mathbb{P}(\mathcal{A} \big| \mathcal{B})$ 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 $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. PMFs are denoted by the capital letter P, with a subscript that identifies the random variable and its possible conditioning. For example, for a discrete probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ 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 $P_{X|Y}(x|y) = \mathbb{P}\big(X = x | Y = y\big)$. We omit subscripts if the

[1] the erasure probability corresponds to the portion of symbols the eavesdropper in the WTC II does not intercept

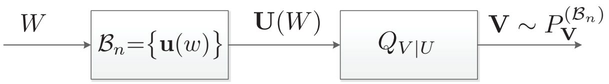
W
$\mathcal{B}_n=\{\mathbf{u}(w)\}$
$\mathbf{U}(W)$
$Q_{V|U}$
$\mathbf{V} \sim P_{\mathbf{V}}^{(\mathcal{B}_n)}$

Fig. 1: Coding problem with the goal of making $P_{\mathbf{V}}^{(\mathcal{B}_n)} \approx Q_V^n$.

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 $\mathbb{E}\big[X\big]$, respectively.

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, $\mathbb{P}_Q\big(\mathcal{A}\big) = \sum_{\omega \in \mathcal{A}} Q(\omega)$, 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 $P_{Y^n|X^n}(\mathbf{y}|\mathbf{x}) = P_{Y|X}^n(\mathbf{y}|\mathbf{x})$. We often use Q_X^n or $Q_{Y|X}^n$ when referring to an i.i.d. sequence of random variables. The conditional product PMF $Q_{Y|X}^n$ given a specific sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted by $Q_{Y|X=\mathbf{x}}^n$.

The empirical PMF $\nu_{\mathbf{x}}$ of a sequence $\mathbf{x} \in \mathcal{X}^n$ is $\nu_{\mathbf{x}}(x) \triangleq \frac{N(x|\mathbf{x})}{n}$, 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 and the non-negative number ϵ [24, Chapter 3], i.e., we have

$$\mathcal{T}_\epsilon^n(P_X) = \Big\{ \mathbf{x} \in \mathcal{X}^n \Big| \ \big|\nu_{\mathbf{x}}(x) - P_X(x)\big| \leq \epsilon P_X(x),\ \forall x \in \mathcal{X} \Big\}. \tag{1}$$

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

$$D(P||Q) = \int_{\mathcal{X}} dP \log\left(\frac{dP}{dQ}\right), \tag{2}$$

where $\frac{dP}{dQ}$ denotes the Radon-Nikodym derivative between P and Q. If the sample space $\mathcal{X}$ is countable, (2) reduces to

$$D(P||Q) = \sum_{x \in \mathrm{supp}(P)} P(x) \log\left(\frac{P(x)}{Q(x)}\right). \tag{3}$$

III. The Stronger Soft-Covering Lemma

We use notation from [25, Section II]. Wyner's soft-covering lemma [13, Theorem 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 $Q_{V|U}$, 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 $\mathcal{B}_n$ 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]. 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 $\mathcal{W} = \big[1 : 2^{nR}\big]$ and $\mathbb{B}_n = \big\{\mathbf{U}(w)\big\}_{w \in \mathcal{W}}$ 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 $\mathcal{B}_n = \big\{\mathbf{u}(w, \mathcal{B}_n)\big\}_{w \in \mathcal{W}}$ denote a realization of $\mathbb{B}_n$. For every fixed $\mathcal{B}_n$, the induced distribution is:

$$P_{\mathbf{V}}^{(\mathcal{B}_n)}(\mathbf{v}) = 2^{-nR} \sum_{w \in \mathcal{W}} Q_{V|U}^n\big(\mathbf{v}\big|\mathbf{u}(w, \mathcal{B}_n)\big). \tag{4}$$

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*

$$\mathbb{P}\bigg(D\Big(P_{\mathbf{V}}^{(\mathbb{B}_n)} \Big|\Big| Q_V^n\Big) > e^{-n\gamma_1} \bigg) \leq e^{-e^{n\gamma_2}}. \tag{5}$$

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

$$\mathbb{P}\bigg(D\Big(P_{\mathbf{V}}^{(\mathbb{B}_n)} \Big|\Big| Q_V^n\Big) > c_\delta n 2^{-n\gamma_\delta} \bigg) \leq \big(1 + |\mathcal{V}|^n\big) e^{-\frac{1}{3}2^{n\delta}}, \tag{6}$$

where

$$\gamma_\delta = \sup_{\alpha > 1} \frac{\alpha - 1}{2\alpha - 1}\big(R - \delta - d_\alpha(Q_{U,V}, Q_U Q_V)\big), \tag{7a}$$

$$c_\delta = 3 \log e + 2\gamma_\delta \log 2 + 2 \log\left(\max_{v \in \mathrm{supp}(Q_V)} \frac{1}{Q_V(v)} \right), \tag{7b}$$

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.

The proof of the lemma is given in Section VI-A.

Remark 1 *The inequality* (6) *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 (6) 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

$$\gamma^* = \sup_{\delta > 0} \gamma_\delta = \gamma_0 = \sup_{\alpha > 1} \frac{\alpha - 1}{2\alpha - 1}\big(R - d_\alpha(Q_{U,V}, Q_U Q_V)\big). \tag{8}$$

The double-exponential confidence rate δ acts as a reduction in codebook rate R in the definition of γ_δ. Consequently, $\gamma_\delta = 0$

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

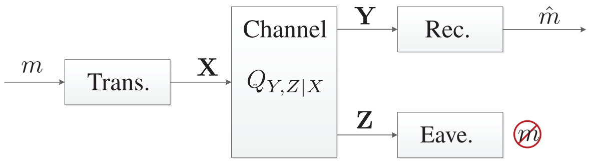

Fig. 2: The classic wiretap channel, referred to as the WTC I.

for $\delta \geq R - I(U;V)$.

IV. Wiretap Channels of Type I

As a rather simple application of stronger soft-covering lemma, we give an alternative derivation of the SS-capacity of the WTC I [3], [20], [26]. 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 and 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 blocklength, the union bound and the double-exponential decay the lemma provides show the existence of a semantically-secure sequence 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.

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) *An (n,R) WTC I code $\mathcal{C}_{1,n}$ has: (i) A message set $\mathcal{M} = \big[1:2^{nR}\big]$; (ii) A stochastic encoder $f_1 : \mathcal{M} \to \mathcal{P}(\mathcal{X}^n)$, where $\mathcal{P}(\mathcal{X}^n)$ denotes the set of all PMFs on $\mathcal{X}^n$; (iii) A decoding function $\phi_1 : \mathcal{Y}^n \to \mathcal{M}$.*

Definition 2 (Maximal Error Probability) *The maximal error probability of an (n,R) WTC II code $\mathcal{C}_{1,n}$ is $e^\star(\mathcal{C}_{1,n}) = \max\limits_{m\in\mathcal{M}} e_m(\mathcal{C}_{n,1})$, where*

$$e_m(\mathcal{C}_{n,1}) = \sum_{\mathbf{x}\in\mathcal{X}^n} f_1(\mathbf{x}|m) \sum_{\substack{\mathbf{y}\in\mathcal{Y}^n:\\ \phi_1(\mathbf{y})\neq m}} Q^n_{Y|X}(\mathbf{y}|\mathbf{x}). \tag{9}$$

Definition 3 (SS Metric) *The SS metric with respect to an $(n,2^{nR})$ WTC I code $\mathcal{C}_{n,1}$ is* [3]

$$\mathrm{Sem}(\mathcal{C}_{n,1}) = \max_{P_M} I_{\mathcal{C}_{n,1}}(M;\mathbf{Z}), \tag{10}$$

where $I_{\mathcal{C}_{n,1}}$ denotes a mutual information term that is calculated with respect to the joint PMF of M and $\mathbf{Z}$ induced by $\mathcal{C}_{n,1}$. Namely, for any $P_M \in \mathcal{P}(\mathcal{M})$, $P^{(\mathcal{C}_{n,1})}_{M,\mathbf{Z}}$ is

$$P^{(\mathcal{C}_{n,1})}_{M,\mathbf{Z}}(m,\mathbf{z}) = P(m) \sum_{\mathbf{x}\in\mathcal{X}^n} f_1(\mathbf{x}|m) \sum_{\mathbf{y}\in\mathcal{Y}^n} Q_{Y,Z|X}(\mathbf{y},\mathbf{z}|\mathbf{x}).$$

Remark 2 *SS requires that a single codebook works well for all message PMF. Accordingly, the maximization over P_M in* (10) *is preformed when the code $\mathcal{C}_{n,1}$ is known. In other words, although not stated explicitly, P_M is a function of $\mathcal{C}_{n,1}$.*

Definition 4 (Semantically-Secure Codes) *A sequence of $(n,2^{nR})$ WTC I codes $\big\{\mathcal{C}_{n,1}\big\}_{n\geq 1}$ is semantically-secure if there is a constants $\gamma > 0$ and an $n_0 \in \mathbb{N}$, such that for every $n > n_0$, $\mathrm{Sem}(\mathcal{C}_{n,1}) \leq e^{-n\gamma}$.*

Remark 3 *By Definition 4, for a sequence of WTC I codes to be semantically-secure, the SS metric from* (10) *must vanish exponentially fast. This is a standard requirement in the cryptography community, commonly referred to as strong-SS (see, e.g., [3, Section 3.2]).*

Definition 5 (SS-Achievability) *A rate $R \in \mathbb{R}_+$ is SS-achievable if there is a sequence of $(n,2^{nR})$ WTC I semantically-secure codes $\big\{\mathcal{C}_{n,1}\big\}_{n\geq 1}$ with $e^\star(\mathcal{C}_{n,1}) \xrightarrow[n\to\infty]{} 0$.*

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 2 (WTC I SS-Capacity) *The SS-capacity of the WTC I is*

$$C_{\mathrm{Sem}} = \max_{\substack{Q_{V,X}:\\ V-X-(Y,Z)}} \Big[I(V;Y) - I(V;Z)\Big], \tag{11}$$

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

The direct proof of Theorem 2 is given in Section VI-B. The converse is straightforward because the weak-secrecy capacity of the WTC I upper bounds its SS-capacity.

V. Wiretap Channels of Type II

A. Problem Definition

The WTC II is illustrated in Fig. 3. The sender chooses a message m from the set $\big[1:2^{nR}\big]$ and maps it into a sequence

[3] $\mathrm{Sem}(\mathcal{C}_{n,1})$ is actually the mutual-information-security metric, which is equivalent to SS by [3]. We use the representation in (10) rather than the formal definition of SS (see, e.g., [3, Equation (4)]) out of analytical convenience.

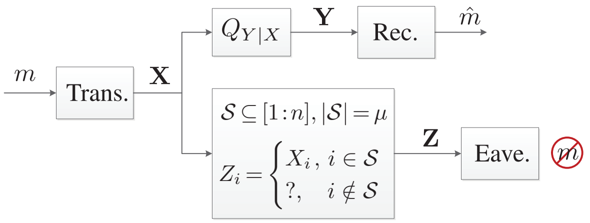
m
Trans.
$\mathbf{X}$
$Q_{Y|X}$
$\mathbf{Y}$
Rec.
$\hat{m}$
$\mathcal{S}\subseteq[1:n], |\mathcal{S}|=\mu$
$Z_i=\begin{cases}X_i, & i\in\mathcal{S}\\ ?, & i\notin\mathcal{S}\end{cases}$
$\mathbf{Z}$
Eave.
$\cancel{m}$

Fig. 3: The type II wiretap channel.

$\mathbf{x}\in\mathcal{X}^n$ (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 $\mathbf{z}\in\big(\mathcal{X}\cup\{?\}\big)^n$, where

$$z_i=\begin{cases}x_i, & i\in\mathcal{S}\\ ?, & i\notin\mathcal{S}\end{cases}. \qquad (12)$$

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

Definition 7 (SS Metric) *The SS metric with respect to a (n,R) WTC II code $\mathcal{C}_{n,2}$ is*

$$\mathrm{Sem}_\mu(\mathcal{C}_{n,2})=\max_{\substack{P_M\in\mathcal{P}(\mathcal{M}),\\ \mathcal{S}\subseteq[1:n]:\ |\mathcal{S}|=\mu}} I_{\mathcal{C}_{n,2}}(M;\mathbf{Z}), \qquad (13)$$

where $I_{\mathcal{C}_{n,2}}$ denotes that the mutual information term is calculated with respect to

$$P^{(\mathcal{C}_{n,2},\mathcal{S})}_{M,\mathbf{Z}}(m,\mathbf{z})=P(m)\sum_{\mathbf{x}\in\mathcal{X}^n}f(\mathbf{x}|m)\mathbb{1}_{\big\{z_i=x_i,\,i\in\mathcal{S}\big\}\cap\big\{z_i=?,\,i\notin\mathcal{S}\big\}}.$$

Remark 4 *Similarly to Remark 2, the mutual information term in* (13) *is maximized over P_M and $\mathcal{S}$ when the code $\mathcal{C}_{n,2}$ is known.*

Definition 8 (Semantically-Secure Codes) *Let $\alpha\in[0,1]$ and $\mu=\lfloor\alpha n\rfloor$, a sequence of (n,R) WTC II codes $\big\{\mathcal{C}_{n,2}\big\}_{n\geq1}$ is α-semantically-secure if there is a constants $\gamma>0$, such that $\mathrm{Sem}_\mu(\mathcal{C}_{n,2})\leq e^{-n\gamma}$ for sufficiently large n.*

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 $\big\{\mathcal{C}_{n,2}\big\}_{n\geq1}$ with $e^\star(\mathcal{C}_{n,2})\xrightarrow[n\to\infty]{}0$.*

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. Capacity 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. Strong-secrecy-capacity is defined with respect to the *average* error probability (instead of the maximal error probability from Definition 2) and the *strong-secrecy* metric (instead of the SS-metric from Definition 7). For example, the strong-secrecy metric of the WTC II is defined similarly to (13) while removing the maximization over $P_M\in\mathcal{P}(\mathcal{M})$ and letting M be a random variable uniformly distributed over $\mathcal{M}$. See [25] for the full definitions.

Proposition 3 (WTC I Upper Bounds WTC II) *Let $\alpha\in[0,1]$ and $C^{\mathrm{II}}_{\mathrm{S}}(\alpha)$ be the α-strong-secrecy-capacity of the WTC II with a main channel $Q^{(2)}_{Y|X}$. Furthermore, let $C^{\mathrm{I}}_{\mathrm{S}}(\alpha)$ be the strong-secrecy-capacity of the WTC I with transition probability $Q^{(1)}_{Y,Z|X}=Q^{(2)}_{Y|X}\mathcal{E}^{(\alpha)}_{Z|X}$, where $\mathcal{E}^{(\alpha)}_{Z|X}$ is a DM-EC with erasure probability $\bar{\alpha}=1-\alpha$, i.e.,*

$$\mathcal{E}^{(\alpha)}_{Z|X}(z|x)=\begin{cases}\alpha, & z=x\\ \bar{\alpha}, & z=?\end{cases}\quad,\ \forall x\in\mathcal{X}. \qquad (14)$$

Then

$$C^{\mathrm{II}}_S(\alpha)\leq C^{\mathrm{I}}_S(\alpha)=\max_{\substack{Q_{V,X}:\\ V-X-Y}}\Big[I(V;Y)-\alpha I(V;X)\Big]. \qquad (15)$$

Due to space limitations, Section VI-C gives only an outline of the proof of Proposition 3 (see [25] for the full derivation). The proof leverages Sanov's Theorem and the continuity of mutual information.

Theorem 4 (WTC II SS-Capacity) *For any $\alpha\in[0,1]$,*

$$C_{\mathrm{Sem}}(\alpha)=\max_{\substack{Q_{V,X}:\\ V-X-Y}}\Big[I(V;Y)-\alpha I(V;X)\Big], \qquad (16)$$

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

The direct part of Theorem 4 is proven in Section VI-D. The stronger soft-covering lemma is key in the security analysis. The converse is a direct consequence of Proposition 3 by noting that the α-strong-secrecy-capacity of the WTC II upper bounds its α-SS-capacity.

Remark 5 *Theorem 4 recovers the achievability result from [8, Equation 7] by setting $V=X$ and taking X to be uniformly distributed over $\mathcal{X}$. Although we require security with respect to a stricter metric than used in [8] (SS versus weak-secrecy), we achieve higher rates than [8, Equation 7] and establish their optimality. Moreover, $C_{\mathrm{Sem}}(\alpha)$ is achieved via classic wiretap codes, making the (rather convoluted) coset coding scheme from [8] (inspired by [7]) no longer required.*

VI. Proofs

A. 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 $\Delta_{\mathcal{B}_n}(\mathbf{v})\triangleq\frac{dP^{(\mathcal{B}_n)}_{\mathbf{V}}}{dQ^n_V}(\mathbf{v})$. In the discrete case, this is just a ratio of probability mass functions. Accordingly, the relative entropy of interest, which is a function of the codebook $\mathcal{B}_n$, is given by

$$D\Big(P_{\mathbf{V}}^{(\mathcal{B}_n)}\Big|\Big|Q_V^n\Big) = \int dP_{\mathbf{V}}^{(\mathcal{B}_n)} \log \Delta_{\mathcal{B}_n}. \tag{17}$$

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

$$i_{Q_{U,V}}(u,v) \triangleq \log\left(\frac{dQ_{V|U=u}}{dQ_V}(v)\right). \tag{18}$$

In (18), the argument of the logarithm is the Radon-Nikodym derivative between $Q_{V|U=u}$ and Q_V. Let $\epsilon \geq 0$ be arbitrary, to be determined later, and define

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

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

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

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

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

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

$$\Delta_{\mathcal{B}_n,j}(\mathbf{v}) \triangleq \frac{dP_{\mathcal{B}_n,j}}{dQ_V^n}(\mathbf{v}), \quad j=1,2. \tag{19}$$

With respect to the above definitions, Lemma 5 states an upper bound on the relative entropy of interest.

Lemma 5 *For every fixed codebook $\mathcal{B}_n$, we have*

$$\begin{aligned} D\Big(P_{\mathbf{V}}^{(\mathcal{B}_n)}\Big|\Big|Q_V^n\Big) &\leq h\left(\int dP_{\mathcal{B}_n,1}\right) \\ &\quad + \int dP_{\mathcal{B}_n,1}\log\Delta_{\mathcal{B}_n,1} + \int dP_{\mathcal{B}_n,2}\log\Delta_{\mathcal{B}_n,2}, \end{aligned} \tag{20}$$

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

Due to space limitations the proof of Lemma 5 is omitted. The reader is referred to [25, Appendix B] for the proof. Based on Lemma 5, 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 (20). 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

$$\begin{aligned} \int dP_{\mathcal{B}_n,2} &= 1 - \int dP_{\mathcal{B}_n,1} \\ &= \sum_{w\in\mathcal{W}} 2^{-nR}\mathbb{P}_{Q_{V|U}^n}\Big(\big(\mathbf{u}(w,\mathcal{B}_n),\mathbf{V}\big)\notin\mathcal{A}_\epsilon\Big|\mathbf{U}=\mathbf{u}(w,\mathcal{B}_n)\Big). \end{aligned} \tag{21}$$

For a random codebook, (21) becomes

$$\begin{aligned} &\int dP_{\mathbb{B}_n,2} \\ &= \sum_{w\in\mathcal{W}} 2^{-nR}\mathbb{P}_{Q_{V|U}^n}\Big(\big(\mathbf{U}(w,\mathbb{B}_n),\mathbf{V}\big)\notin\mathcal{A}_\epsilon\Big|\mathbf{U}=\mathbf{U}(w,\mathbb{B}_n)\Big). \end{aligned} \tag{22}$$

The RHS of (22) 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:

$$\begin{aligned} &\mathbb{E}_{\mathbb{B}_n}\mathbb{P}_{Q_{V|U}^n}\Big(\big(\mathbf{U}(w,\mathbb{B}_n),\mathbf{V}\big)\notin\mathcal{A}_\epsilon\Big|\mathbf{U}=\mathbf{U}(w,\mathbb{B}_n)\Big) \\ &\qquad = \mathbb{P}_{Q_{U,V}^n}\Big(\big(\mathbf{U},\mathbf{V}\big)\notin\mathcal{A}_\epsilon\Big) \\ &\qquad \overset{(a)}{=} \mathbb{P}_{Q_{U,V}^n}\left(2^{\lambda\sum_{t=1}^n i_{Q_{U,V}}(U_t,V_t)} \geq 2^{n\lambda(I(U;V)+\epsilon)}\right) \\ &\qquad \overset{(b)}{\leq} \frac{\mathbb{E}_{Q_{U,V}^n} 2^{\lambda\sum_{t=1}^n i_{Q_{U,V}}(U_t,V_t)}}{2^{n\lambda(I(U;V)+\epsilon)}} \\ &\qquad = \left(\frac{\mathbb{E}_{Q_{U,V}} 2^{\lambda i_{Q_{U,V}}(U,V)}}{2^{\lambda(I(U;V)+\epsilon)}}\right)^n \\ &\qquad \overset{(c)}{=} 2^{n\lambda\left(\frac{1}{\lambda}\log_2\mathbb{E}_{Q_{U,V}}\left[2^{\lambda i_{Q_{U,V}}(U;V)}\right]-I(U;V)-\epsilon\right)} \\ &\qquad \overset{(d)}{=} 2^{n\lambda\left(d_{\lambda+1}(Q_{U,V},Q_UQ_V)-I(U;V)-\epsilon\right)}, \end{aligned} \tag{23}$$

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 (23) gives

$$\mathbb{E}_{\mathbb{B}_n}\mathbb{P}_{Q_{V|U}^n}\Big(\big(\mathbf{U}(w,\mathbb{B}_n),\mathbf{V}\big)\notin\mathcal{A}_\epsilon\Big|\mathbf{U}=\mathbf{U}(w,\mathbb{B}_n)\Big) \leq 2^{-n\beta_{\alpha,\epsilon}}, \tag{24}$$

where $\beta_{\alpha,\epsilon} = (\alpha-1)\big(I(U;V)+\epsilon-d_\alpha(Q_{U,V},Q_UQ_V)\big)$ 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 theorem statement, and set

$$\epsilon_{\alpha,\delta} = \frac{\frac{1}{2}(R-\delta)+(\alpha-1)d_\alpha(Q_{U,V},Q_UQ_V)}{\frac{1}{2}+(\alpha-1)} - I(U;V). \tag{25}$$

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

$$\beta_{\alpha,\delta} \triangleq \beta_{\alpha,\epsilon_{\alpha,\delta}} = \frac{\alpha-1}{2\alpha-1}\big(R-\delta-d_\alpha(Q_{U,V},Q_UQ_V)\big). \tag{26}$$

Observe that $\epsilon_{\alpha,\delta}$ in (25) is nonnegative under the assumption that $R-\delta > I(U;V)$, because $\alpha>1$ and $d_\alpha(Q_{U,V},Q_UQ_V) \geq d_1(Q_{U,V},Q_UQ_V) = I(U;V)$.

Next, we use the following version of the Chernoff bound to bound the probability of (22) not being exponentially small.

Lemma 6 (Chernoff Bound [25]) *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 $\frac{c}{\mu} \in [1,2]$,*

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M} X_m \geq c\right) \leq e^{-\frac{M\mu}{3B}\left(\frac{c}{\mu}-1\right)^2}. \tag{27}$$

Using (27) 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

$$\mathbb{P}\left(\int dP_{\mathcal{B}_n,2} \geq 2 \cdot 2^{-n\beta_{\alpha,\delta}}\right) \leq e^{-\frac{1}{3}2^{n(R-\beta_{\alpha,\delta})}}. \tag{28}$$

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:

$$\begin{aligned}&\Delta_{\mathbb{B}_n,1}(\mathbf{v})\\ &= \sum_{w\in\mathcal{W}} 2^{-nR}\frac{dQ^n_{V|U=\mathbf{U}(w,\mathbb{B}_n)}}{dQ^n_V}(\mathbf{v})\mathbf{1}_{\left\{\left(\mathbf{U}(w,\mathbb{B}_n),\mathbf{v}\right)\in\mathcal{A}_\epsilon\right\}}.\end{aligned} \tag{29}$$

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 (27), gives

$$\begin{aligned}\mathbb{P}\Big(\Delta_{\mathbb{B}_n,1}(\mathbf{v}) \geq 1 + 2^{-n\beta_{\alpha,\delta}}\Big) &\leq e^{-\frac{1}{3}2^{n(R-I(U;V)-\epsilon_{\alpha,\delta}-2\beta_{\alpha,\delta})}}\\ &= e^{-\frac{1}{3}2^{n\delta}}, \quad \forall\, \mathbf{v} \in \mathcal{V}^n,\end{aligned} \tag{30}$$

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

$$\Delta_{\mathbb{B}_n,2}(\mathbf{v}) \leq \left(\max_{v\in\text{supp}(Q_V)} \frac{1}{Q_V(v)}\right)^n, \ \forall\, \mathbf{v} \in \mathcal{V}^n, \tag{31}$$

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 (28), (30) and (31), we can now bound the probability that the RHS of (20) is not exponentially small. Let $\mathcal{S}$ be the set of codebooks $\mathcal{B}_n$, such that all of the following are true:

$$\int dP_{\mathcal{B}_n,2} < 2 \cdot 2^{-n\beta_{\alpha,\delta}}, \tag{32a}$$

$$\Delta_{\mathcal{B}_n,1}(\mathbf{v}) < 1 + 2^{-n\beta_{\alpha,\delta}}, \ \forall\, \mathbf{v} \in \mathcal{V}^n, \tag{32b}$$

$$\Delta_{\mathcal{B}_n,2}(\mathbf{v}) \leq \left(\max_{v\in\text{supp}(Q_V)} \frac{1}{Q_V(v)}\right)^n, \ \forall\, \mathbf{v} \in \mathcal{V}^n. \tag{32c}$$

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:

$$\begin{aligned}\mathbb{P}\big(\mathbb{B}_n \notin \mathcal{S}\big) &\overset{(a)}{\leq} \mathbb{P}\left(\int dP_{\mathbb{B}_n,2} \geq 2 \cdot 2^{-n\beta_{\alpha,\delta}}\right)\\ &\quad + \sum_{\mathbf{v}\in\mathcal{V}^n}\mathbb{P}\Big(\Delta_{\mathbb{B}_n,1}(\mathbf{v}) \geq 1 + 2^{-\beta_{\alpha,\delta}n}\Big)\\ &\quad + \sum_{\mathbf{v}\in\mathcal{V}^n}\mathbb{P}\left(\Delta_{\mathbb{B}_n,2}(\mathbf{v}) > \left(\max_{v\in\text{supp}(Q_V)}\frac{1}{Q_V(v)}\right)^n\right)\\ &\overset{(b)}{\leq} e^{-\frac{1}{3}2^{n(R-\beta_{\alpha,\delta})}} + |\mathcal{V}|^n \cdot e^{-\frac{1}{3}2^{n\delta}}\\ &\overset{(c)}{\leq} \left(1 + |\mathcal{V}|^n\right)e^{-\frac{1}{3}2^{n\delta}},\end{aligned} \tag{33}$$

where (a) is the union bound, (b) uses (28), (30) and (31), 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 (20) 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 (32a) implies that

$$h\left(\int dP_{\mathcal{B}_n,2}\right) < 2\big(\log e + \beta_{\alpha,\delta}\log 2\big)n2^{-n\beta_{\alpha,\delta}}. \tag{34}$$

Furthermore, by (32b) and since $\log(1+x) \leq x\log e$, for every $x > 0$, we have

$$\int dP_{\mathcal{B}_n,1}\log\Delta_{\mathcal{B}_n,1} < 2^{-n\beta_{\alpha,\delta}}\log e. \tag{35}$$

Finally, using (32a) and (32c) we obtain

$$\int dP_{\mathcal{B}_n,2}\log\Delta_{\mathcal{B}_n,2} < 2\log\left(\max_{v\in\text{supp}(Q_V)}\frac{1}{Q_V(v)}\right)n2^{-n\beta_{\alpha,\delta}}. \tag{36}$$

Combining (34)-(36), yields

$$\begin{aligned}h\left(\int dP_{\mathcal{B}_n,1}\right) + \int dP_{\mathcal{B}_n,1}\log\Delta_{\mathcal{B}_n,1} + \int dP_{\mathcal{B}_n,2}\log\Delta_{\mathcal{B}_n,2}\\ \overset{(a)}{<} c_{\alpha,\delta}n2^{-n\beta_{\alpha,\delta}},\end{aligned} \tag{37}$$

where (a) comes from setting

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

Through Lemma 5, the above implies that for all $\alpha > 1$ and $\delta \in \big(0, R - I(U;V)\big)$,

$$\begin{aligned}\mathbb{P}\left(D\Big(P^{(\mathbb{B}_n)}_{\mathbf{V}}\Big|\Big|Q^n_V\Big) \geq c_{\alpha,\delta}n2^{-n\beta_{\alpha,\delta}}\right) &\leq \mathbb{P}\big(\mathbb{B}_n \notin \mathcal{S}\big)\\ &\overset{(a)}{\leq} \left(1 + |\mathcal{V}|^n\right)e^{-\frac{1}{3}2^{n\delta}},\end{aligned} \tag{38}$$

where (a) follows from (33). Denoting $c_\delta \triangleq \sup_{\alpha>1} c_{\alpha,\delta}$, (38) further gives

$$\mathbb{P}\Big(D\Big(P^{(\mathbb{B}_n)}_{\mathbf{V}}\Big|\Big|Q^n_V\Big) \geq c_\delta n 2^{-n\beta_{\alpha,\delta}}\Big) \leq \left(1+|\mathcal{V}|^n\right) e^{-\frac{1}{3}2^{n\delta}}. \tag{39}$$

Since (39) is true for all $\alpha > 1$, it must also be true, with strict inequality in the LHS, when replacing $\beta_{\alpha,\delta}$ with

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

which is the exponential rate of convergence stated in (7a) that we derive for the strong soft-covering lemma. This establishes the statement from (6) 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

$$\mathbb{P}\Big(D\Big(P^{(\mathbb{B}_n)}_{\mathbf{V}}\Big|\Big|Q^n_V\Big) > e^{-n\gamma_1}\Big) \leq e^{-e^{n\gamma_2}}. \tag{40}$$

B. Direct Proof of Theorem 2

We show the achievability of (11) when $V = X$. Then, a standard channel prefixing argument extends the proof to any V with $V - X - Y$. Furthermore, we present here a construction of a sequence of semantically-secure codes with a vanishing *average* error probability. The expurgation technique [21, Theorem 7.7.1] allows upgrading reliability to achieve a vanishing *maximal* error probability, while preserving SS.

Fix $\epsilon > 0$ and a PMF $Q_X \in \mathcal{P}(\mathcal{X})$, and let M and W be independent random variables uniformly distributed over $\mathcal{M}$ and $\mathcal{W} = \big[1 : 2^{n\tilde{R}}\big]$, respectively.

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

Encoder f_1: To send $m \in \mathcal{M}$ the encoder randomly and uniformly chooses w from $\mathcal{W}$ and transmits $\mathbf{x}(m,w,\mathcal{B}_n)$.

Decoder ϕ_1: Upon observing $\mathbf{y}$, the receiver searches for a unique pair $(\hat{m},\hat{w}) \in \mathcal{M}\times\mathcal{W}$ such that $\big(\mathbf{x}(\hat{m},\hat{w},\mathcal{B}_n),\mathbf{y}\big) \in \mathcal{T}^n_\epsilon(Q_{X,Y})$. If such a unique pair is found, then $\hat{m}$ is declared as the decoded message; otherwise, an error is declared.

The triple $(\mathcal{M}, f_1, \phi_1)$ defined with respect to the codebook $\mathcal{B}_n$ constitute an (n,R) code for the WTC II. When a random codebook $\mathbb{B}_n$ is used, we denote the corresponding random code by $\mathbb{C}_n$. Standard joint-typicality decoding arguments show that

$$\mathbb{E}_{\mathbb{C}_n} \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} e_m(\mathbb{C}_n) \xrightarrow[n\to\infty]{} 0, \tag{41}$$

(see (9)) provided that

$$R + \tilde{R} < I(X;Y). \tag{42}$$

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

$$\begin{aligned} I_{\mathcal{C}_n}(M;\mathbf{Z}) &= D\left(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M}\middle|\middle|P^{(\mathcal{C}_n)}_{\mathbf{Z}}\middle|P_M\right)\\ &= D\left(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M}\middle|\middle|Q^n_Z\middle|P_M\right) - D\left(P^{(\mathcal{C}_n)}_{\mathbf{Z}}\middle|\middle|Q^n_Z\middle|P_M\right), \end{aligned} \tag{43}$$

where Q_Z is the marginal of $Q_X Q_{Y,Z|X}$. Therefore

$$\begin{aligned} &\max_{P_M\in\mathcal{P}(\mathcal{M})} I_{\mathcal{C}_n}(M;\mathbf{Z}) \leq \max_{P_M\in\mathcal{P}(\mathcal{M})} D\left(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M}\middle|\middle|Q^n_Z\middle|P_M\right)\\ &\leq \max_{P_M\in\mathcal{P}(\mathcal{M})} \sum_{m\in\mathcal{M}} P(m) \max_{\tilde{m}\in\mathcal{M}} D\left(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M=\tilde{m}}\middle|\middle|Q^n_Z\right)\\ &= \max_{m\in\mathcal{M}} D\left(P^{(\mathcal{C}_n)}_{\mathbf{Z}|M=m}\middle|\middle|Q^n_Z\right). \end{aligned} \tag{44}$$

Now, for an arbitrary $\tilde{\gamma} > 0$ (to be determined later) consider

$$\begin{aligned} &\mathbb{P}\bigg(\Big\{\mathrm{Sem}(\mathbb{C}_n) \leq e^{-n\tilde{\gamma}}\Big\}^c\bigg)\\ &\overset{(a)}{\leq} \mathbb{P}\left(\max_{m\in\mathcal{M}} D\left(P^{(\mathbb{C}_n)}_{\mathbf{Z}|M=m}\middle|\middle|Q^n_Z\right) > e^{-n\tilde{\gamma}}\right)\\ &\overset{(b)}{\leq} \sum_{m\in\mathcal{M}} \mathbb{P}\bigg(D\left(P^{(\mathbb{C}_n)}_{\mathbf{Z}|M=m}\middle|\middle|Q^n_Z\right) > e^{-n\tilde{\gamma}}\bigg), \end{aligned} \tag{45}$$

where (a) uses (44), and (b) is the union bound.

By the stronger soft-covering lemma, if

$$\tilde{R} > I(X;Z), \tag{46}$$

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

$$\mathbb{P}\bigg(D\left(P^{(\mathbb{C}_n)}_{\mathbf{Z}|M=m}\middle|\middle|Q^n_Z\right) > e^{-n\tilde{\gamma}}\bigg) \leq e^{-e^{n\gamma_2}}, \tag{47}$$

for sufficiently large n. Inserting (47) into (45) while setting $\tilde{\gamma} = \gamma_1$, we have

$$\mathbb{P}\bigg(\Big\{\mathrm{Sem}(\mathbb{C}_n) \leq e^{-n\gamma_1}\Big\}^c\bigg) \leq 2^{nR}\cdot e^{-e^{n\gamma_2}} \xrightarrow[n\to\infty]{} 0. \tag{48}$$

Inequality (48) implies that if $\tilde{R}$ satisfies (46), the probability that a randomly generated sequence of codes meets the SS criterion for large n is arbitrarily close to 1. In fact, because (48) decays so rapidly, the Borel-Cantelli lemma implies that almost every sequence of realizations of $\big\{\mathbb{C}_n\big\}_{n\geq1}$ is semantically-secure. Having (41) and (48), the selection lemma [25, Lemma 5] implies that there exists a sequence of (n,R) semantically-secure codes $\big\{\mathcal{C}_{n,1}\big\}_{n\geq1}$ with a vanishing average error probability.

The final step is to amend $\big\{\mathcal{C}_{n,1}\big\}_{n\geq1}$ to be reliable with respect to the maximal error probability (cf. Definition 2). This is done using the standard expurgation technique (see, e.g., [21, Theorem 7.7.1]). Namely, we discard the worst half of the codewords in each $\mathcal{C}_{n,1}$. Denoting the amended sequence of codes by $\big\{\mathcal{C}^\star_{n,1}\big\}_{n\geq1}$, we have $e^\star(\mathcal{C}^\star_{n,1}) \xrightarrow[n\to\infty]{} 0$. To conclude the proof, note that in each $\mathcal{C}^\star_{n,1}$ 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 $\big\{\mathcal{C}_{n,1}\big\}_{n\geq1}$ is semantically-secure, so is $\big\{\mathcal{C}^\star_{n,1}\big\}_{n\geq1}$.

Combining (42) with (46), we have that every

$$0 \le R < \max_{Q_X} \Big[I(X;Y) - I(X;Z) \Big] \tag{49}$$

is SS-achievable. If the RHS of (49) is non-positive, then we set $R = 0$.

To establish the achievability of C_{Sem} from (11), we prefix a DM-channel (DMC) $Q_{X|V}$ to the original WTC I $Q_{Y,Z|X}$ to obtain a new channel $Q_{Y,Z|V}$, where

$$Q^n_{Y,Z|V}(\mathbf{y},\mathbf{z}|\mathbf{v}) = \sum_{\mathbf{x}\in\mathcal{X}^n} Q^n_{X|V}(\mathbf{x}|\mathbf{v}) Q^n_{Y,Z|X}(\mathbf{y},\mathbf{z}|\mathbf{x}). \tag{50}$$

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

$$R < \max_{\substack{Q_{V,X}:\\ V-X-(Y,Z)}} \Big[I(V;Y) - I(V;Z) \Big] \tag{51}$$

is achievable.

C. Proof Outline of Proposition 3

The equality in (15) follows by evaluating the strong-secrecy-capacity formula of a general WTC I, i.e.,

$$\max_{\substack{Q_{V,X}:\\ V-X-(Y,Z)}} \Big[I(V;Y) - I(V;Z) \Big], \tag{52}$$

under the transition probability matrix $Q^{(1)}_{Y,Z|X}$.

To prove the inequality in (15), we first 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 $\bar{\beta}$. Having this, Proposition 3 follows by a continuity argument of the strong-secrecy-capacity of WTC I with respect to β.

For any α-strong-secrecy-achievable rate $R \in \mathbb{R}_+$ for the WTC II, let $\big\{\mathcal{C}^{\mathrm{II}}_n\big\}_{n\ge1}$ be the corresponding sequence of (n,R) codes. Since the channels to the legitimate receivers are the same for both versions of the WTC, $\big\{\mathcal{C}^{\mathrm{II}}_n\big\}_{n\ge1}$ also achieves reliability for the WTC I.

Denote by $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ the eavesdroppers observation in the WTC I and the WTC II, respectively, that is induced by $\big\{\mathcal{C}^{\mathrm{II}}_n\big\}_{n\ge1}$. To establish (15), it suffices to show that $I_{\mathcal{C}^{\mathrm{II}}_n}\big(M;\mathbf{Z}^{(1)}\big)$ can be made arbitrarily small with the blocklength n. To this end, for every $\mathbf{z} \in \mathcal{Z}^n$, where $\mathcal{Z} \triangleq \mathcal{X} \cup \{?\}$, define

$$\mathcal{A}(\mathbf{z}) \triangleq \big\{ i \in [1:n] \big| z_i = ? \big\} \tag{53}$$

$$\Theta(\mathbf{Z}) \triangleq \mathbb{1}_{\big\{|\mathcal{A}(\mathbf{Z})| \le \lceil \bar{\alpha} n \rceil\big\}}. \tag{54}$$

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 $I_{\mathcal{C}^{\mathrm{II}}_n}\big(M;\mathbf{Z}^{(1)}\big)$ on $\Theta(\mathbf{Z}^{(1)})$, we distinguish between the two cases of $\mathbf{Z}^{(1)}$ being better or worse than $\mathbf{Z}^{(2)}$. 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 $\big\{\mathcal{C}^{\mathrm{II}}_n\big\}_{n\ge1}$ achieve security for the WTC II. Otherwise, for the case that $\Theta(\mathbf{Z}^{(1)}) = 1$ and $\mathbf{Z}^{(1)}$ 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.

D. Direct Proof of Theorem 4

As before, we start by showing the achievability of (16) when $V = X$. After doing so, we use channel prefixing to extend the proof for any V with $V - X - Y$.

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 $\mathcal{W} = \big[1 : 2^{n\tilde{R}}\big]$, respectively, we repeat the code construction from Section VI-B and set $f_2 = f_1$ and $\phi_2 = \phi_1$. A similar analysis of the average error probability shows that if

$$R + \tilde{R} < I(X;Y), \tag{55}$$

then (41) holds for the WTC II scenario as well.

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

$$\Gamma^{(\mathcal{S})}_{\mathbf{Z}}(\mathbf{z}) = \prod_{j\in\mathcal{S}^c} \mathbb{1}_{\big\{z_j = ?\big\}} \prod_{j\in\mathcal{S}} \mathcal{I}_Z(z_j), \quad \forall \mathbf{z} \in \mathcal{Z}^n, \tag{56}$$

where $\mathcal{I}_Z$ is the marginal output PMF of the identity DMC with input distribution Q_X:

$$\mathcal{I}_Z(z) = \sum_{x\in\mathcal{X}} Q_X(x) \mathbb{1}_{\{z=x\}} = Q_X(z) \mathbb{1}_{\{z\in\mathcal{X}\}}. \tag{57}$$

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

$$\begin{aligned} I_{\mathcal{C}_n}(M;\mathbf{Z}) &= D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}|M} \middle|\middle| P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}} \middle| P_M \right) \\ &= D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}|M} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}} \middle| P_M \right) - D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}} \middle| P_M \right), \end{aligned} \tag{58}$$

and therefore

$$\begin{aligned} \max_{P_M\in\mathcal{P}(\mathcal{M})} I_{\mathcal{C}_n}(M;\mathbf{Z}) &\le \max_{P_M\in\mathcal{P}(\mathcal{M})} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}|M} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}} \middle| P_M \right) \\ &\le \max_{P_M\in\mathcal{P}(\mathcal{M})} \sum_{m\in\mathcal{M}} P(m) \max_{\tilde{m}\in\mathcal{M}} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}|M=\tilde{m}} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}} \right) \\ &= \max_{m\in\mathcal{M}} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}|M=m} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}} \right). \end{aligned} \tag{59}$$

For any $\emptyset \neq \mathcal{A} \subseteq [1:n]$ and $\mathbf{z} \in \mathcal{Z}^n$, let $\mathbf{z}^{\mathcal{A}} \triangleq (z_i)_{i\in\mathcal{A}}$ be the sub-vector of $\mathbf{z}$ indexed by the elements of $\mathcal{A}$. The relative entropy chain rule further simplifies the RHS of (59) as follows. For any $m \in \mathcal{M}$, we have

$$\begin{aligned} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}|M=m} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}} \right) &= D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}},\mathbf{Z}^{\mathcal{S}^c}|M=m} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}},\mathbf{Z}^{\mathcal{S}^c}} \right) \\ &\overset{(a)}{=} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}|M=m} \middle|\middle| \Gamma^{(\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}} \right) \\ &\overset{(b)}{=} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}|M=m} \middle|\middle| \mathcal{I}^{\mu}_{Z} \right), \end{aligned} \tag{60}$$

where (a) is because $P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}^c}|M=m,\mathbf{Z}^{\mathcal{S}}=\mathbf{z}^{\mathcal{S}}} = \mathbb{1}_{\big\{Z_i=?,\ i\in\mathcal{S}^c\big\}} = \Gamma^{(\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}^c}}$, for every $\mathbf{z}^{\mathcal{S}} \in \mathcal{Z}^{|\mathcal{S}|}$, and (b) follows from (56).

Combining (58)-(60) while maximizing over all subsets $\mathcal{S}$, for every $\mathcal{C}_n$ we have

$$\mathrm{Sem}_\mu(\mathcal{C}_n) \le \max_{\substack{m\in\mathcal{M},\\ \mathcal{S}\subseteq[1:n]:\ |\mathcal{S}|=\mu}} D\left(P^{(\mathcal{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}|M=m} \middle|\middle| \mathcal{I}^{\mu}_{Z} \right). \tag{61}$$

Now, for an arbitrary $\tilde{\gamma} > 0$ (to be determined later) consider

$$\begin{aligned}
&\mathbb{P}\bigg(\Big\{\mathrm{Sem}_\mu(\mathbb{C}_n) \leq e^{-n\tilde{\gamma}}\Big\}^c\bigg)\\
&\overset{(a)}{\leq} \mathbb{P}\left(\max_{\substack{m\in\mathcal{M},\\ \mathcal{S}\subseteq[1:n]:\ |\mathcal{S}|=\mu}} D\left(P^{(\mathbb{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}|M=m}\middle|\middle|\mathcal{I}^\mu_Z\right) > e^{-n\tilde{\gamma}}\right)\\
&\overset{(b)}{\leq} \sum_{\substack{m\in\mathcal{M},\\ \mathcal{S}\subseteq[1:n]:\ |\mathcal{S}|=\mu}} \mathbb{P}\bigg(D\left(P^{(\mathbb{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}|M=m}\middle|\middle|\mathcal{I}^\mu_Z\right) > e^{-n\tilde{\gamma}}\bigg), \qquad (62)
\end{aligned}$$

where (a) uses (61), and (b) is the union bound.

Each term in the sum on the RHS of (62) 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

$$\frac{n\tilde{R}}{\mu} > H(X), \qquad (63)$$

there exist $\gamma_1, \gamma_2 > 0$ that for sufficiently large n satisfy

$$\mathbb{P}\bigg(D\left(P^{(\mathbb{C}_n,\mathcal{S})}_{\mathbf{Z}^{\mathcal{S}}|M=m}\middle|\middle|\mathcal{I}^\mu_Z\right) > e^{-n\gamma_1}\bigg) \leq e^{-e^{n\gamma_2}}. \qquad (64)$$

Since $\mu = \lfloor\alpha n\rfloor \leq \alpha n$, taking $\tilde{R} > \alpha H(X)$ satisfies (63). Setting $\tilde{\gamma} = \gamma_1$ into (62) and using (64), gives

$$\mathbb{P}\Big(\mathrm{Sem}_\mu(\mathbb{C}_n) \geq e^{-n\gamma_1}\Big) \leq 2^n \cdot 2^{nR} \cdot e^{-e^{n\gamma_2}} \xrightarrow[n\to\infty]{} 0. \qquad (65)$$

Having (41) and (65), the selection lemma [25, Lemma 5] implies the existance of a sequence of (n,R) α-semantically-secure codes $\big\{\mathcal{C}_{n,2}\big\}_{n\geq 1}$ with a vanishing average error probability. The expurgation technique again upgrades reliability to be with respect to the maximal error probability while preserving SS. Combining both rate bounds shows the achievability of

$$R < \max_{Q_X}\Big[I(X;Y) - \alpha H(X)\Big]. \qquad (66)$$

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

$$Q^n_{Y|V}(\mathbf{y}|\mathbf{v}) = \sum_{\mathbf{x}\in\mathcal{X}^n} Q^n_{X|V}(\mathbf{x}|\mathbf{v}) Q^n_{Y|X}(\mathbf{y}|\mathbf{x}). \qquad (67)$$

Furthermore, $\Gamma^{(\mathcal{S})}_{\mathbf{Z}}$ from (56) is redefined as

$$\Gamma^{(\mathcal{S})}_{\mathbf{Z}}(\mathbf{z}) = \prod_{j\in\mathcal{S}^c} \mathbb{1}_{\{z_j=?\}} \prod_{j\in\mathcal{S}} Q_Z(z_j), \quad \forall \mathbf{z}\in\mathcal{Z}^n, \qquad (68)$$

where Q_Z is given by

$$Q_Z(z) = \begin{cases} \sum_{v\in\mathcal{V}} Q_V(v) Q_{X|V}(z|v), & z\in\mathcal{X}\\ 0, & z=? \end{cases}. \qquad (69)$$

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

$$R + \tilde{R} < I(V;Y), \qquad (70)$$

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

$$\tilde{R} > \alpha I(V;X). \qquad (71)$$

Putting (70)-(71) together completes the proof of Theorem 4.

Acknowledgment

The work of Z. Goldfeld and H. H. Permuter was supported by the Israel Science Foundation (grant no. 684/11), an ERC starting grant and the Cyber Security Research Grant at Ben-Gurion University of the Negev. The work of P. Cuff was supported by the National Science Foundation (grant CCF-1350595) and the Air Force Office of Scientific Research (grant FA9550-15-1-0180).

References

[1] A. D. Wyner. The wire-tap channel. *Bell Sys. Techn.*, 54(8):1355–1387, Oct. 1975.

[2] M. Bellare and S. Tessaro. Polynomial-time, semantically-secure encryption achieving the secrecy capacity. *Cryptology ePrint Archive*, Report 2012/022, 2012. Available at http://eprint.iacr.org/.

[3] M. Bellare, S. Tessaro, and A. Vardy. A cryptographic treatment of the wiretap channe. In *Proc. Adv. Crypto. (CRYPTO 2012)*, Santa Barbara, CA, USA, Aug. 2012.

[4] S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Comp. and Sys. Sci.*, 28(2):270–299, Apr. 1984.

[5] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A concrete security treatment of symmetric encryption. In *Proc. IEEE 38th Symp. Foundations of Comp. Sci.*, pages 394–403, Miami, Florida, US, Oct. 1997.

[6] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348, May 1978.

[7] L. Ozarow and A. D. Wyner. Wire-tap channel II. *Bell Sys. Techn. Journal*, 63(10):2135–2157, Dec. 1984.

[8] M. Nafea and A. Yener. Wiretap channel II with a noisy main channel. In *Proc. Int. Symp. Inf. Theory (ISIT-2015)*, Hong-Kong, Jun. 2015.

[9] M. J. Mihaljević. On message protection in cryptosystems modeled as the generalized wire-tap channel II. In *Lecture Notes in Computer Science*, pages 13–24, Berlin, Germany, 1994. Springer-Verlag.

[10] Y. Luo, C. Mitrpant, and A. J. H. Vinck. Some new characters on the wire-tap channel of type II. *IEEE Trans. Inf. Theory*, 51(3):1222–, Mar. 2005.

[11] R. Liu, Y. Liang, and H. V. Poor. Secure nested codes for type II wiretap channels. In *Proc. Inf. Theory Workshop (ITW-2007)*, Lake Tahoe, California, US, Sep. 2007.

[12] V. Aggarwal, L. Lai A. R. Calderbank, and H. V. Poor. Wiretap channel type II with an active eavesdropper. In *Proc. Int. Symp. Inf. Theory (ISIT-2009)*, Seoul, Korea, Jun.-Jul. 2009.

[13] A. D. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2):163–179, Mar. 1975.

[14] M. Bloch and N. Laneman. Strong secrecy from channel resolvability. *IEEE Trans. Inf. Theory*, 59(12):8077–8098, Dec. 2013.

[15] C. Schieler and P. Cuff. The henchman problem: Measuring secrecy by the minimum distortion in a list. *submitted to IEEE Trans. Inf. Theory*, 2014.

[16] C. Schieler and P. Cuff. Rate-distortion theory for secrecy systems. *IEEE Trans. Inf. Theory*, 66(12):7584–7605, Dec. 2014.

[17] T. Han and S.Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–772, May 1993.

[18] P. Cuff. Distributed channel synthesis. *IEEE. Trans. Inf. Theory*, 59(11):7071–7096, Nov. 2013.

[19] E. Song, P. Cuff, and V. Poor. The likelihood encoder for lossy compression. *submitted to IEEE Trans. Inf. Theory*, Aug. 2014.

[20] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé. Semantically secure lattice codes for the Gaussian wiretap channel. *IEEE Trans. Inf. Theory*, 60(10):6399–6416, 2014.

[21] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[22] V. K. Wei. Generalized Hamming weights for linear codes. *IEEE Trans. Inf. Theory*, 37(5):1412–1418, Sep. 1991.

[23] 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*, 53(8):2933–2945, Aug. 2007.

[24] J. L. Massey. *Applied Digital Information Theory*. ETH Zurich, Zurich, Switzerland, 1980-1998.

[25] Z. Goldfeld, P. Cuff, and H. H. Permuter. Semantic-security capacity for wiretap channels of type II. *Submitted to IEEE Trans. Inf. Theory*, 2015. Available on ArXiv at http://arxiv.org/abs/1509.03619.

[26] M. Hayashi and R. Matsumoto. Secure multiplex coding with dependent and non-uniform multiple messages. *IEEE Trans. Inf. Theory*, 2016. Accepted for publication.