

A New Wiretap Channel Model and Its Strong Secrecy Capacity

Mohamed Nafea, *Student Member, IEEE*, and Aylin Yener^{ID}, *Fellow, IEEE*

Abstract—In this paper, a new wiretap channel model is proposed, where the legitimate transmitter and receiver communicate over a discrete memoryless channel. The wiretapper has perfect access to a fixed-length subset of the transmitted code word symbols of her choosing. Additionally, she observes the remainder of the transmitted symbols through a discrete memoryless channel. This new model subsumes the classical wiretap channel and wiretap channel II with noisy main channel as its special cases, and is termed as the generalized wiretap channel for that reason. The strong secrecy capacity of the proposed channel model is identified. Achievability is established by solving a dual secret key agreement problem in the source model, and converting the solution to the original channel model using probability distribution approximation arguments. In the dual problem, a source encoder and decoder, who observe random sequences independent and identically distributed according to the input and output distributions of the legitimate channel in the original problem, communicate a confidential key over a public error-free channel using a single forward transmission, in the presence of a compound wiretapping source who has perfect access to the public discussion. The security of the key is guaranteed for the exponentially many possibilities of the subset chosen at the wiretapper by deriving a lemma which provides a *doubly-exponential* convergence rate for the probability that, for a fixed choice of the subset, the key is uniform and independent from the public discussion and the wiretapping source’s observation. The converse is derived by using Sanov’s theorem to upper bound the secrecy capacity of the generalized wiretap channel by the secrecy capacity when the tapped subset is randomly chosen by nature.

Index Terms—Wiretap channel, wiretap channel II, strong secrecy, source-channel duality, random binning, concentration inequalities.

I. INTRODUCTION

WYNER’S wiretap channel models a legitimate transmitter and a receiver communicating over a discrete memoryless channel (DMC), referred to as the main channel, in the presence of a passive wiretapper who *only listens* to the transmitted signal through a cascaded second DMC, referred to as the wiretapper channel [2]. Subsequently, reference [3] has generalized Wyner’s wiretap channel model to a general, not necessarily degraded, discrete memoryless wiretap channel. Later, Ozarow and Wyner, in reference [4], have introduced

the wiretap channel II model, which considers a noiseless main channel and a binary erasure channel to the wiretapper, where the wiretapper is able to select the positions of erasures. Interestingly, using random partitioning and combinatorial arguments, reference [4] has showed that the secrecy capacity for this channel is equal to that if the wiretapper is a passive observer unable to choose the positions of the erasures, thus demonstrating the ability of coding to neutralize this more powerful wiretapper.

While considerable research on code design for secure communication followed the randomized coset coding of [4], see for example [5]–[7], the idea of the wiretap channel II remained linked to the assumption of a noiseless main channel for several decades, mainly due to technical challenges in generalizing the model outside of this special model. Yet, the notion of providing the wiretapper with this additional capability of choosing what to observe is appealing and represents a positive step towards providing confidentiality guarantees in stronger attack models. Towards this end, reference [8] introduced a discrete memoryless (noisy) main channel to the wiretap channel II model, and derived outer and inner bounds for the capacity-equivocation region of the model, where the proposed achievability scheme is optimal for the special case of the maximizing input distribution being uniform. More recently, reference [9] found the secrecy capacity of this model, showing that, once again, the secrecy capacity is equal to that of the case when the wiretapper channel is replaced with a discrete memoryless erasure channel.

This work goes one step further and introduces a *new wiretap channel model* with a discrete memoryless main channel and a wiretapper who observes a subset of the transmitted codeword symbols of her choosing perfectly, as well as observing the remaining symbols through a second DMC. This model includes as special cases both the classical wiretap channel in [3] by setting the subset size to zero, and the wiretap channel II with a noisy main channel in [8] by setting the wiretapper’s DMC to an erasure channel with erasure probability one, and is termed as the *generalized wiretap channel* for that reason. We characterize the *strong* secrecy capacity for the proposed wiretap channel model, quantifying precisely the cost in secrecy capacity due to the additional capability at the wiretapper, with respect to the previous wiretap models.

We first present the achievability. Recent independent work [9] has provided an achievability proof for the wiretap channel II model with a noisy main channel considered in [8] using a stronger version of Wyner’s soft covering

Manuscript received October 23, 2016; revised August 14, 2017; accepted November 12, 2017. Date of publication December 22, 2017; date of current version February 15, 2018. This work was supported by NSF under Grant CCF 13-19338 and Grant CNS 13-14719. This paper was presented in part at the 2016 IEEE International Symposium on Information Theory [1].

The authors are with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA 16802 USA (e-mail: mnafea@psu.edu; yener@ee.psu.edu).

Communicated by H. Permuter, Associate Editor for Shannon Theory. Digital Object Identifier 10.1109/TIT.2017.2786541

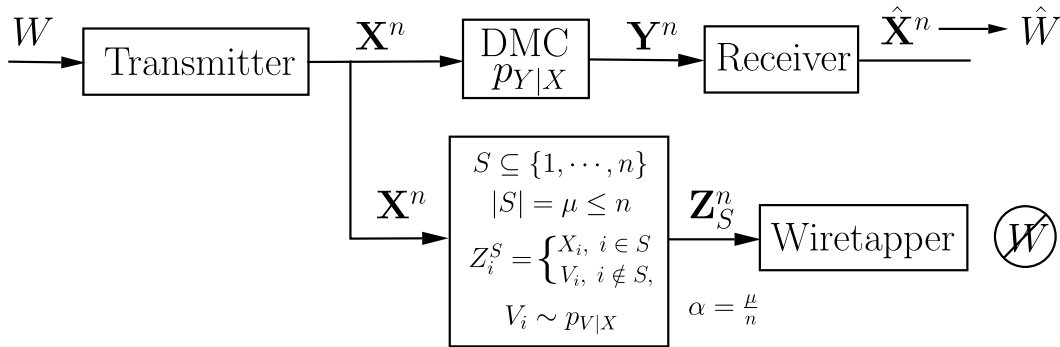


Fig. 1. The generalized wiretap channel: A new wiretap channel model.

lemma [10]. Although, with a careful handling, this approach can be applied to the wiretap channel model we propose in the paper, we follow an alternative approach to establish its achievability, namely by using a framework similar to the output statistics of random binning framework in [11]. In particular, we solve a dual secret key agreement problem in the source model sense [12], [13], and infer the design for the encoder and decoder of the original channel model from the solution of the dual problem. The difference between our achievability proof and the framework presented in [11] is that we measure the statistical dependence between the transmitted message and the wiretapper's observation in terms of the Kullback-Leibler (K-L) divergence instead of total variation distance, which requires establishing a convergence result, with a rate strictly faster than $\frac{1}{n}$, for the probability that the two induced distributions from the original and the dual models are close in the total variation distance sense. In addition, in the source model, we guarantee the secrecy of the confidential key for the exponentially many possibilities of the subset chosen at the wiretapper by deriving a *one-shot* result which provides a *doubly-exponential convergence rate* for the probability that the key is uniform and independent from the wiretapper's observation.

The converse is derived by identifying a channel model whose secrecy capacity is identical to that of the proposed channel model, and is easier to establish the converse of. This is done by means of upper bounding its secrecy capacity with that of a discrete memoryless channel¹ whose secrecy capacity is tractable.

The remainder of the paper is organized as follows. Section II describes the new wiretap channel model. Section III provides the main result of the paper, i.e., the strong secrecy capacity for the new wiretap channel. Sections IV and V provide the achievability and converse proofs. Section VI provides a discussion about the main result and the adopted achievability approach. Section VII concludes the paper. The proofs for the supporting lemmas are provided in the Appendices.

II. CHANNEL MODEL AND DEFINITIONS

We first remark the notation we use throughout the paper. Vectors are denoted by bold lower-case super-scripted letters while their components are denoted by lower-case sub-scripted letters. A similar convention but with upper-case letters is used

for random vectors and their components. Vector superscripts are omitted when dimensions are clear from the context. We use $\mathbb{1}\{\mathcal{A}\}$ to denote the indicator function of the event \mathcal{A} . For $a, b \in \mathbb{R}$, $\llbracket a, b \rrbracket$ denotes the set of integers $\{i \in \mathbb{N} : a \leq i \leq b\}$. For $S \subseteq \mathbb{N}$, \mathbf{X}_S denotes the sequence $\{X_i\}_{i \in S}$. We use upper-case letters to denote random probability distributions, e.g., P_X , and lower-case letters to denote deterministic probability distributions, e.g., p_X . We use p_X^U to denote a uniform distribution over the random variable X . The argument of the probability distribution is omitted when it is clear from its subscript. $\mathbb{V}(p_X, q_X)$ and $\mathbb{D}(p_X || q_X)$ denote the total variation distance and the Kullback-Leibler (K-L) divergence between the probability distributions p_X and q_X .

We consider the channel model illustrated in Figure 1. The main channel $\{\mathcal{X}, \mathcal{Y}, p_{Y|X}\}$ is a discrete memoryless channel (DMC) which consists of a finite input alphabet \mathcal{X} , a finite output alphabet \mathcal{Y} , and a transition probability $p_{Y|X}$. The transmitter wishes to transmit a message W , uniformly distributed over $\mathcal{W} = \llbracket 1, 2^{nR_s} \rrbracket$, to the legitimate receiver reliably, and to keep the message secret from the wiretapper. To do so, the transmitter maps the message W to the transmitted codeword $\mathbf{X}^n \in \mathcal{X}^n$ using a stochastic encoder. The legitimate receiver observes $\mathbf{Y}^n \in \mathcal{Y}^n$ and maps its observation to the estimate \hat{W} of the message W . The wiretapper chooses a subset $S \in \mathcal{S}$ where the set \mathcal{S} is defined as

$$\mathcal{S} \triangleq \left\{ S : S \subseteq \llbracket 1, n \rrbracket, |S| = \mu \leq n, \alpha = \frac{\mu}{n} \right\}. \quad (1)$$

Then, the wiretapper observes the sequence $\mathbf{Z}_S^n \triangleq [Z_1^S, Z_2^S, \dots, Z_n^S] \in \mathcal{Z}^n$, with

$$Z_i^S = \begin{cases} X_i, & i \in S \\ V_i, & \text{otherwise,} \end{cases} \quad (2)$$

where $\mathbf{V}^n \triangleq [V_1, V_2, \dots, V_n] \in \mathcal{V}^n$ is the output of the DMC $p_{V|X}$ when \mathbf{X}^n is the input, and the alphabet \mathcal{Z} is given by $\mathcal{Z} = \{\mathcal{X} \cup \mathcal{V}\}$.

An $(n, 2^{nR_s})$ code \mathcal{C}_n for the channel model in Figure 1 consists of

- (i) the message set $\mathcal{W} = \llbracket 1, 2^{nR_s} \rrbracket$,
- (ii) the stochastic encoder $P_{\mathbf{X}^n|W, \mathcal{C}_n}$ at the transmitter, and
- (iii) the decoder at the legitimate receiver.

We consider the strong secrecy constraint at the wiretapper [14], [15]. Rate R_s is an achievable strong secrecy rate if there exists a sequence of $(n, 2^{nR_s})$ channel

¹A similar approach was considered to derive [9, Proposition 1].

codes, $\{\mathcal{C}_n\}_{n \geq 1}$, such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\hat{W} \neq W | \mathcal{C}_n \right) = 0 \quad \text{Reliability,} \quad (3)$$

$$\text{and } \lim_{n \rightarrow \infty} \max_{S \in \mathcal{S}} I(W; \mathbf{Z}_S^n | \mathcal{C}_n) = 0 \quad \text{Strong Secrecy,} \quad (4)$$

where \mathcal{S} is defined as in (1). The strong secrecy capacity, C_s , is the supremum of all achievable strong secrecy rates.

Finally, we will be using the following two measures extensively in the sequel. The total variation distance between two probability distributions p_X and q_X , defined on the same probability space, is given by

$$\begin{aligned} \mathbb{V}(p_X, q_X) &\triangleq \frac{1}{2} \sum_{x \in \mathcal{X}} |p(x) - q(x)| \\ &= \sum_{x \in \mathcal{X}: p(x) > q(x)} [p(x) - q(x)]. \end{aligned} \quad (5)$$

The Kullback-Leibler divergence, or relative entropy, between the two distributions p_X and q_X , defined on the same probability space, is given by

$$\mathbb{D}(p_X || q_X) \triangleq \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}. \quad (6)$$

III. MAIN RESULT

The main result of this paper is stated in the following theorem.

Theorem 1: For $0 \leq \alpha \leq 1$, the strong secrecy capacity of the generalized wiretap channel in Figure 1 is given by

$$C_s(\alpha) = \max_{U-X-YV} [I(U; Y) - I(U; V) - \alpha I(U; X|V)]^+, \quad (7)$$

where the maximization is over all the distributions p_{UX} which satisfy the Markov chain $U - X - YV$, and the cardinality of U can be restricted as $|\mathcal{U}| \leq |\mathcal{X}|$.

Proof: The achievability and converse proofs for Theorem 1 are provided in Sections IV and V, respectively. ■

Remark 1: An equivalent characterization for the strong secrecy capacity of the generalized wiretap channel is given by

$$C_s(\alpha) = \max_{U-X-YV} \left[\begin{array}{c} I(U; Y) - \alpha I(U; X) \\ -(1 - \alpha) I(U; V) \end{array} \right]^+, \quad (8)$$

since $I(U; X|V)$ in (7) can be written as

$$I(U; X|V) = H(U|V) - H(U|X) \quad (9)$$

$$= H(U) - I(U; V) - H(U|X) \quad (10)$$

$$= I(U; X) - I(U; V), \quad (11)$$

where (9) follows from the Markov chain $U - X - V$.

Corollary 1: By setting the tapped subset by the wiretapper, S , to the null set, or equivalently $\alpha = 0$, the secrecy capacity in (7) is equal to the secrecy capacity of the discrete memoryless wiretap channel in [3, Corollary 2], i.e.,

$$C_s(0) = \max_{U-X-YV} [I(U; Y) - I(U; V)]^+. \quad (12)$$

Remark 2: Comparing (7) and (12), we observe that the secrecy cost, with respect to the classical wiretap channel, of the additional capability of the wiretapper to choose a subset

of size αn of the codewords to access perfectly, is equal to $\alpha I(U; X|V)$.

Corollary 2: By setting the wiretapper's DMC through which she observes the $(1 - \alpha)n$ symbols she does not choose, $p_{V|X}$, to be an erasure channel with erasure probability one, the secrecy capacity in (7) is equal to the secrecy capacity of the wiretap channel II with a noisy main channel in [9, Th. 2], i.e.,

$$C_s(\alpha) = \max_{U-X-Y} [I(U; Y) - \alpha I(U; X)]^+. \quad (13)$$

Remark 3: Comparing (8) and (13), the secrecy cost, with respect to the wiretap channel II with a noisy main channel, of the additional capability of the wiretapper of observing $(1 - \alpha)$ fraction of the codeword through the DMC $p_{V|X}$, is equal to $(1 - \alpha)I(U; V)$.

IV. ACHIEVABILITY

We establish the achievability for Theorem 1 using an indirect approach as in [11], [16], and [17]. We first assume the availability of a certain common randomness at all terminals of the original channel model. We then define a *dual* secret key agreement problem in the source model which introduces a set of random variables similar to the set of variables introduced by the original problem with the assumed common randomness. The alphabets of the random variables in the original and dual problems are identical. In addition, a subset of the marginal and conditional distributions for these random variables in the original and dual problems are considered to be identical. Yet, the joint distribution of the random variables in the dual problem can differ from that of the original problem due to the different dynamics in the two problems. The main trick is to search for conditions such that the joint distributions of the random variables in the two problems are almost identical in the total variation distance sense. This enables converting the solution, i.e., finding an encoder and decoder which satisfy certain reliability and secrecy conditions, for the dual problem, which is more tractable, to a solution of the original problem. We finally eliminate the assumed common randomness from the original channel model by conditioning on a certain instance of it. Duality here is an *operational* duality [18] in which the solution for the dual problem is converted to a solution for the original problem.

We first prove the achievability for the case $U = X$. We fix the input distribution p_X and define two protocols; each of these protocols introduces a set of random variables and random vectors and induces a joint distribution over them. The first protocol, protocol A, describes a dual secret key agreement problem in which a source encoder and decoder observe random sequences independently and identically distributed (i.i.d.) according to the input and output distributions of the original channel model. The source encoder and decoder intend to communicate a confidential key via transmitting a public message over an error-free channel, in the presence of a *compound* wiretapping source who has perfect access to the public message and observes another random sequence whose distribution belongs to a finite class of distributions, with no prior distribution over the class. The second protocol,

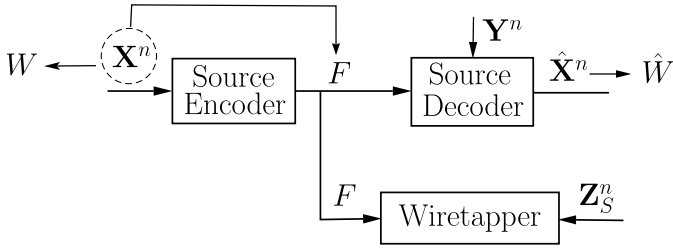


Fig. 2. Protocol A: Secret key agreement in the source model.

protocol B, describes the original channel model in Figure 1, with the addition of assuming a common randomness that is available at all terminals. In the following, we describe the two protocols in detail.

Protocol A (Secret Key Agreement in Source Model): The protocol is illustrated in Figure 2. The random vectors $\mathbf{X}^n, \mathbf{Y}^n$ are i.i.d. according to $p_{XY} = p_X p_{Y|X}$, where $p_{Y|X}$ is the transition probability of the main channel in Figure 1. The source encoder observes the sequence \mathbf{X}^n and randomly assigns (bins) it into the two bin indices $W = \mathcal{B}_{1,n}(\mathbf{X}^n)$ and $F = \mathcal{B}_{2,n}(\mathbf{X}^n)$, where $\mathcal{B}_{1,n}$ and $\mathcal{B}_{2,n}$ are uniformly distributed over $\llbracket 1, 2^{nR_s} \rrbracket$ and $\llbracket 1, 2^{n\tilde{R}_s} \rrbracket$, respectively. That is, each $\mathbf{x}^n \in \mathcal{X}^n$ is randomly and independently assigned to the indices $w \in \llbracket 1, 2^{nR_s} \rrbracket$ and $f \in \llbracket 1, 2^{n\tilde{R}_s} \rrbracket$. The bin index F represents the public message which is transmitted over a noiseless channel to the decoder and perfectly accessed by the wiretapper. The bin index W represents the confidential key to be generated at the encoder and reconstructed at the decoder. The source decoder observes F and the i.i.d. sequence \mathbf{Y}^n , and outputs the estimate $\hat{\mathbf{X}}^n$ of \mathbf{X}^n , which in turn generates the estimate \hat{W} of W . For any $S \in \mathcal{S}$, where \mathcal{S} is defined as in (1), the wiretapper source node observes F and the sequence \mathbf{Z}_S^n in (2). The subset S is selected by the wiretapper and her selection is unknown to the legitimate parties. Thus, the wiretapper can be represented as a compound source $\mathbf{Z}_S^n \triangleq \{z, p_{Z_S^n}, S \in \mathcal{S}\}$ whose distribution is only known to belong to the finite class $\{p_{Z_S^n}\}_{S \in \mathcal{S}}$ with no prior distribution over the class, with $|\mathcal{S}| = \binom{n}{an} \leq 2^n$. For $S \in \mathcal{S}$, the induced joint distribution for this protocol is

$$\begin{aligned} & \tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) \\ &= p_{XYZ_S}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \tilde{P}_{WF|X}(w, f|\mathbf{x}) \tilde{P}_{\hat{\mathbf{X}}|YF}(\hat{\mathbf{x}}|\mathbf{y}, f) \end{aligned} \quad (14)$$

$$\begin{aligned} &= p_{XYZ_S}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \mathbb{1}\{\mathcal{B}_{1,n}(\mathbf{X}) = W\} \mathbb{1}\{\mathcal{B}_{2,n}(\mathbf{X}) = F\} \\ &\quad \times \tilde{P}_{\hat{\mathbf{X}}|YF}(\hat{\mathbf{x}}|\mathbf{y}, f) \end{aligned} \quad (15)$$

$$\begin{aligned} &= \tilde{P}_{WF}(w, f) \tilde{P}_{X|WF}(\mathbf{x}|w, f) p_{YZ_S|X}(\mathbf{y}, \mathbf{z}|\mathbf{x}) \\ &\quad \times \tilde{P}_{\hat{\mathbf{X}}|YF}(\hat{\mathbf{x}}|\mathbf{y}, f). \end{aligned} \quad (16)$$

Protocol B (Main problem assisted with common randomness): This protocol is defined as the channel model in Figure 1, with an addition of a common randomness F that is uniformly distributed over $\llbracket 1, 2^{n\tilde{R}_s} \rrbracket$, independent from all other variables, and known at all terminals. In fact, the assumed common randomness represents the random nature in generating the codebook, which is known at all nodes. At the end of the proof, we eliminate the assumed common randomness from the channel model in this protocol by conditioning on a certain instance of it. The encoder and decoder in this protocol are defined as in (16),

i.e., $P_{X|WF} = \tilde{P}_{X|WF}$ and $P_{\hat{X}|YF} = \tilde{P}_{\hat{X}|YF}$. The induced joint distribution for this protocol is given by

$$\begin{aligned} & P_{WFXYZ_S \hat{\mathbf{X}}}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) \\ &= p_W^U p_F^U \tilde{P}_{X|WF}(\mathbf{x}|w, f) p_{YZ_S|X}(\mathbf{y}, \mathbf{z}|\mathbf{x}) \tilde{P}_{\hat{\mathbf{X}}|YF}(\hat{\mathbf{x}}|\mathbf{y}, f). \end{aligned} \quad (17)$$

The induced joint distributions in (16) and (17) are random due to the random binning of \mathbf{X}^n . Note that we have ignored the random variables \hat{W} from the induced joint distributions at this stage. We will introduce them later to the joint distributions as deterministic functions of the random vectors $\hat{\mathbf{X}}^n$, after fixing the binning functions.

The remaining steps of the proof are outlined as follows:

- (i) We derive a condition on the rates R_s and \tilde{R}_s such that the two induced joint distributions (16) and (17) are close in the total variation distance sense, when averaged over the random binning.
- (ii) We then use Slepian-Wolf source coding theorem [19], [20] to derive a condition on the rate \tilde{R}_s such that the decoding of $\hat{\mathbf{X}}$ in protocol A is reliable, i.e., the communication of the key W is reliable.
- (iii) Next, for protocol A, we derive another condition on the rates R_s and \tilde{R}_s such that the probability, with respect to the random binning, that for all $S \in \mathcal{S}$, the key W and the public message F are uniformly distributed, independent, and both independent from the wiretapper's observation \mathbf{Z}_S^n , goes to one as n goes to infinity, i.e., protocol A is secure.
- (iv) We use the closeness of the two induced distributions for the two protocols to show that, under the same rate conditions for protocol B, the aforementioned reliability and secrecy properties in (ii) and (iii) hold for protocol B as well.
- (v) The reliability and secrecy properties in (ii) and (iii), after being converted to the channel model in protocol B, are averaged over the random binning of the dual source model² in protocol A. We show the existence of a fixed binning realization such that both properties still hold for protocol B.
- (vi) Finally, we eliminate the common randomness F from the channel model in protocol B by showing that the reliability and secrecy constraints still hold when we condition on a certain instance of F , i.e., $F = f^*$.

Note that, for the secrecy constraint, we have required the independence of the assumed common randomness F from both W and \mathbf{Z}_S^n so that when we condition over an instance of F , the independence of W and \mathbf{Z}_S^n is not affected. That is, the secrecy (independence) property in (iii) for protocol B, after fixing the binning function and removing the common randomness, results in an achievable strong secrecy rate for the original channel model. Before continuing with the proof, we state the following lemmas.

A. Useful Lemmas

Lemma 1 is a *one-shot* result, which provides an exponential decay rate for the average, over the random binning, of the

²Note that the probability with respect to the random binning in (iii) is equivalent to an average over the random binning of an indicator function.

total variation distance between the two induced distributions from the two protocols. We utilize this lemma to show a convergence in probability result that allows converting the secrecy property from protocol A to protocol B. A result similar to Lemma 1 was derived in [11, Appendix A] which does not provide the required convergence rate, hence the need for Lemma 1.

Lemma 1: Let the source $X \triangleq \{\mathcal{X}, p_X\}$ be randomly binned into $W = \mathcal{B}_1(X)$ and $F = \mathcal{B}_2(X)$, where \mathcal{B}_1 and \mathcal{B}_2 are uniform over $\llbracket 1, \tilde{W} \rrbracket$ and $\llbracket 1, \tilde{F} \rrbracket$, respectively. Let $\mathcal{B} \triangleq \{\mathcal{B}_1(x), \mathcal{B}_2(x)\}_{x \in \mathcal{X}}$, and for $\gamma > 0$, define

$$\mathcal{D}_\gamma \triangleq \left\{ x \in \mathcal{X} : \log \frac{1}{p_X(x)} > \gamma \right\}. \quad (18)$$

Then, we have

$$\mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(P_{WF}, P_W^U P_F^U \right) \right) \leq \mathbb{P}(X \notin \mathcal{D}_\gamma) + \frac{1}{2} \sqrt{\tilde{W} \tilde{F} 2^{-\gamma}}, \quad (19)$$

where P is the induced distribution over W and F .

Proof: The proof is provided in Appendix A. ■

Lemma 2 below is again a one-shot result which provides a *doubly-exponential* decay rate for the probability of failure of achieving the secrecy property for protocol A, for a fixed choice of the subset S . This lemma is needed, along with the union bound, to guarantee secrecy against the exponentially many possibilities of the tapped subset S .

Lemma 2: Let $X \triangleq \{\mathcal{X}, p_X\}$ and $\{Z_S\} \triangleq \{Z, p_{Z_S}, S \in \mathcal{S}\}$ be two correlated sources with $|\mathcal{X}|, |\mathcal{Z}|$, and $|\mathcal{S}| < \infty$, where $\{Z_S\}_{S \in \mathcal{S}}$ is a compound source whose distribution is known to belong to the finite class $\{p_{Z_S}\}_{S \in \mathcal{S}}$. Let X be randomly binned into the bin indices W and F as in Lemma 1. For $\gamma > 0$ and any $S \in \mathcal{S}$, define

$$\mathcal{D}_\gamma^S \triangleq \left\{ (x, z) \in \mathcal{X} \times \mathcal{Z} : \log \frac{1}{p_{X|Z_S}(x|z)} > \gamma \right\}. \quad (20)$$

If there exists $\delta \in (0, \frac{1}{2})$ such that for all $S \in \mathcal{S}$, $\mathbb{P}_{p_{XZ_S}} \left((X, Z_S) \in \mathcal{D}_\gamma^S \right) \geq 1 - \delta^2$, then, we have, for every $\epsilon_1 \in [0, 1]$, that

$$\mathbb{P}_{\mathcal{B}} \left(\max_{S \in \mathcal{S}} \mathbb{D}(P_{WFZ_S} \| P_W^U P_F^U p_{Z_S}) \geq \tilde{\epsilon} \right) \leq |\mathcal{S}| |\mathcal{Z}| \exp \left(\frac{-\epsilon_1^2 (1 - \delta) 2^\gamma}{3 \tilde{W} \tilde{F}} \right), \quad (21)$$

where $\tilde{\epsilon} = \epsilon_1 + (\delta + \delta^2) \log(\tilde{W} \tilde{F}) + H_b(\delta^2)$, H_b is the binary entropy function, and P is the induced distribution over W, F , and Z_S .

Proof: The proof of Lemma 2 is given in Appendix B. ■

The selection lemma below is used to show the existence of a binning realization such that both the secrecy and reliability properties hold for protocol B. It is also used to eliminate the common randomness F from the channel model in protocol B.

Lemma 3 (Selection Lemma [21, Lemma 2.2]): Let A_1, A_2, \dots, A_n be a sequence of random variables where $A_n \in \mathcal{A}^n$, and let $\mathcal{F}_n \triangleq \{f_{1,n}, \dots, f_{M,n}\}$ be a finite set of

bounded functions $f_{i,n} : \mathcal{A}^n \mapsto \mathbb{R}^+$, $i \in \llbracket 1, M \rrbracket$, such that $|\mathcal{F}_n| = M$ does not depend on n , and

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{A}^n} (f_{i,n}(A_n)) = 0 \quad \text{for all } i \in \llbracket 1, M \rrbracket. \quad (22)$$

Then, there exists a specific realization $\{a_n^*\}$ of the sequence $\{A_n\}$ such that

$$\lim_{n \rightarrow \infty} f_{i,n}(a_n^*) = 0 \quad \text{for all } i \in \llbracket 1, M \rrbracket. \quad (23)$$

The following Lemma states two properties of the total variation distance, which we utilize through the achievability proof.

Lemma 4 (Properties of Total Variation Distance [17, Lemmas V.1 and V.2]): Consider the joint distributions $p_{XY} = p_X p_{Y|X}$ and $q_{XY} = q_X q_{Y|X}$, defined on the same probability space. Then, we have,

$$\mathbb{V}(p_X, q_X) \leq \mathbb{V}(p_{XY}, q_{XY}) \quad (24)$$

$$\mathbb{V}(p_X p_{Y|X}, q_X p_{Y|X}) = \mathbb{V}(p_X, q_X). \quad (25)$$

In order to apply Lemmas 1 and 2 to protocol A, we use Hoeffding's inequality, which is stated in the following Lemma.

Lemma 5 (Hoeffding's Inequality [22, Th. 2]): Let X_1, X_2, \dots, X_n be independent random variables with $X_i \in [0, b]$ for all $i \in \llbracket 1, n \rrbracket$, and let $\bar{m} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i)$. Then, for $\epsilon > 0$, we have

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n X_i \leq (1 - \epsilon) \bar{m} \right) \leq \exp \left(\frac{-2\epsilon^2 \bar{m}^2}{b^2} n \right). \quad (26)$$

B. Proof

First, we apply Lemma 1 to protocol A. In Lemma 1, set $X = \mathbf{X}^n$, $\tilde{W} = 2^{nR_s}$, $\tilde{F} = 2^{n\tilde{R}_s}$, $\mathcal{B} = \mathcal{B}_n \triangleq \{\mathcal{B}_{1,n}(\mathbf{x}), \mathcal{B}_{2,n}(\mathbf{x})\}_{\mathbf{x} \in \mathcal{X}^n}$, and $\gamma = n(1 - \epsilon_2)H(X)$, where $\epsilon_2 > 0$ and \mathbf{X}^n is defined as in protocol A, i.e., is an i.i.d. sequence. Without loss of generality, we assume that for all $x \in \mathcal{X}$, we have $p_X(x) > 0$. Let $p_{\min} = \min_{x \in \mathcal{X}} p_X(x)$, where the minimum exists since the input alphabet \mathcal{X} is finite.³ Thus, the random variables $\log \frac{1}{p_X(X_i)}$, $i \in \llbracket 1, n \rrbracket$, are i.i.d. and each is bounded by the interval $[0, b_{\max}]$, where $b_{\max} = \log \frac{1}{p_{\min}}$.

We also have that $\bar{m} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{p_X} \left(\log \frac{1}{p_X(X_i)} \right) = H(X)$. Using Hoeffding's inequality in (26), we have, for any $\epsilon_2 > 0$, that

$$\mathbb{P}(\mathbf{X} \notin \mathcal{D}_\gamma) = \mathbb{P}_{p_X} \left(\log \frac{1}{p_X(\mathbf{X})} \leq \gamma \right) \quad (27)$$

$$= \mathbb{P}_{p_X} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{1}{p_X(X_i)} \leq (1 - \epsilon_2) H(X) \right) \quad (28)$$

$$\leq \exp \left(\frac{-2\epsilon_2^2 H^2(X)}{b_{\max}^2} n \right) = \exp(-\beta_1 n), \quad (29)$$

where $\beta_1 = \frac{2\epsilon_2^2 H^2(X)}{b_{\max}^2} > 0$.

³If the input alphabet \mathcal{X} is infinite, $\min_{x \in \mathcal{X}} p_X(x)$ might not exist. As a result, there might not be a finite upper bound on the random variables $\log \frac{1}{p_X(X_i)}$. In such a case, Hoeffding inequality can not be applied.

By substituting the choices for \tilde{W} , \tilde{F} , γ and (29) in (19), we have, as long as $R_s + \tilde{R}_s < (1 - \epsilon_2)H(X)$, that

$$\mathbb{E}_{\mathcal{B}_n}(\mathbb{V}(\tilde{P}_{WF}, p_{WF}^U p_F^U)) \leq 2 \exp(-\beta n), \quad (30)$$

where $\beta_2 = \frac{\ln 2}{2} \left((1 - \epsilon_2)H(X) - R_s - \tilde{R}_s \right)$ and $\beta = \min\{\beta_1, \beta_2\} > 0$. By applying (25) to (16) and (17), and using (30), we have

$$\begin{aligned} & \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{V} \left(\tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}, P_{WFXYZ_S \hat{\mathbf{X}}} \right) \right) \\ &= \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{V} \left(\tilde{P}_{WF}, p_{WF}^U p_F^U \right) \right) \leq 2 \exp(-\beta n). \end{aligned} \quad (31)$$

Consider Slepian-Wolf decoder for protocol A. As long as $\tilde{R}_s \geq H(X|Y)$, we have [23, Th. 10.1]

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{P}_{\tilde{P}}(\hat{\mathbf{X}} \neq \mathbf{X}) \right) = 0. \quad (32)$$

Next, we observe

$$\begin{aligned} & \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{V} \left(\tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \right) \\ &= \mathbb{E}_{\mathcal{B}_n} \sum_{\substack{w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}: \\ \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}) \mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\} \\ < \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}})}} \left[\begin{array}{l} \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) \\ - \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}) \mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\} \end{array} \right] \end{aligned} \quad (33)$$

$$= \mathbb{E}_{\mathcal{B}_n} \sum_{w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}: \hat{\mathbf{x}} \neq \mathbf{x}} \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) \quad (34)$$

$$= \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{P}_{\tilde{P}}(\hat{\mathbf{X}} \neq \mathbf{X}) \right). \quad (35)$$

Equation (34) follows because $\tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) > \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}) \mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\}$ holds if and only if $\mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\} = 0$, where $\tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}})$ factorizes as $\tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) = \tilde{P}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}) \tilde{P}(\hat{\mathbf{x}}|\mathbf{y}, f)$ and $\tilde{P}(\hat{\mathbf{x}}|\mathbf{y}, f) \leq 1$. Thus, using (32) and (35), we have that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{V} \left(\tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \right) = 0, \quad (36)$$

as long as $\tilde{R}_s \geq H(X|Y)$.

Now, we apply Lemma 2 to protocol A. In Lemma 2, set $X = \mathbf{X}^n$, $\tilde{W} = 2^{nR_s}$, $\tilde{F} = 2^{n\tilde{R}_s}$, $\mathcal{B} = \mathcal{B}_n$, $Z_S = \mathbf{Z}_S^n$, for all $S \in \mathcal{S}$, and $\gamma = n(1 - \tilde{\epsilon}_2)(1 - \alpha)H(X|V)$, where $\tilde{\epsilon}_2 > 0$ and $\mathbf{X}^n, \mathbf{Z}_S^n, \mathcal{S}$ are defined as in protocol A. In order to calculate $\mathbb{P}_{p_{\mathbf{XZ}_S}}(\mathbf{X}, \mathbf{Z}_S) \notin \mathcal{D}_\gamma^S$, we only need to consider the pairs (\mathbf{x}, \mathbf{z}) such that $p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) > 0$, since all the pairs (\mathbf{x}, \mathbf{z}) with $p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) = 0$ belong to \mathcal{D}_γ^S , by the definition of \mathcal{D}_γ^S in (20). Since the sequence \mathbf{X} is i.i.d. and the channel $p_{V|X}$ is memoryless, we have, for all (\mathbf{x}, \mathbf{z}) with $p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) > 0$, that

$$p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) = p_{\mathbf{X}_S \mathbf{X}_{S^c} | \mathbf{X}_S \mathbf{V}_{S^c}}(\mathbf{x}_S, \mathbf{x}_{S^c} | \mathbf{x}_S, \mathbf{v}_{S^c}) \quad (37)$$

$$= p_{\mathbf{X}_{S^c} | \mathbf{V}_{S^c}}(\mathbf{x}_{S^c} | \mathbf{v}_{S^c}) = \prod_{i \in S^c} p_{X|V}(x_i | v_i). \quad (38)$$

Once again, using Hoeffding's inequality, we have, for all $S \in \mathcal{S}$,

$$\begin{aligned} & \mathbb{P}_{p_{\mathbf{XZ}_S}} \left((\mathbf{X}, \mathbf{Z}_S) \notin \mathcal{D}_\gamma^S \right) \\ &= \mathbb{P}_{p_{\mathbf{XZ}_S}} \left(p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{X}|\mathbf{Z}_S) > 0, \log \frac{1}{p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{X}|\mathbf{Z}_S)} \leq \gamma \right) \end{aligned} \quad (39)$$

$$= \mathbb{P}_{p_{X|V}} \left(\frac{1}{n - \mu} \sum_{i \in S^c} \log \frac{1}{p_{X|V}(X_i | V_i)} \leq (1 - \tilde{\epsilon}_2)H(X|V) \right) \quad (40)$$

$$\leq \exp(-\tilde{\beta}(1 - \alpha)n) = \delta^2, \quad (41)$$

where $\tilde{\beta} > 0$, and (40) follows from (38). From (41), $\lim_{n \rightarrow \infty} \delta^2 = 0$, and hence, for sufficiently large n , we have $\delta^2 \in (0, \frac{1}{4})$. Thus, the conditions in Lemma 2 are satisfied.

Note that $\lim_{n \rightarrow \infty} n(\delta + \delta^2) = 0$, and $\lim_{n \rightarrow \infty} H_b(\delta^2) = H_b(\lim_{n \rightarrow \infty} \delta^2) = 0$ since H_b is a continuous function. Thus,

$$\lim_{n \rightarrow \infty} \tilde{\epsilon} = \epsilon_1 + (R_s + \tilde{R}_s) \lim_{n \rightarrow \infty} n(\delta + \delta^2) + \lim_{n \rightarrow \infty} H_b(\delta^2) = \epsilon_1. \quad (42)$$

By substituting the choices for \tilde{W} , \tilde{F} , γ , and $|\mathcal{S}||\mathcal{Z}^n| \leq \exp(n[\ln 2 + \ln(|\mathcal{X}| + |\mathcal{V}|)])$ in (21), and using (42), we have that, for all $\epsilon_1, \epsilon'_1 > 0$ and $\tilde{\epsilon} = \epsilon_1 + \epsilon'_1$, there exist $n^* \in \mathbb{N}$ and $\psi(\epsilon_1), \kappa > 0$ such that, for all $n \geq n^*$,

$$\begin{aligned} & \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(\tilde{P}_{WFZ_S} \| p_W^U p_F^U p_{Z_S} \right) \geq \tilde{\epsilon} \right) \\ & \leq \exp(-\psi(\epsilon_1)e^{\kappa n}), \end{aligned} \quad (43)$$

as long as $R_s + \tilde{R}_s < (1 - \tilde{\epsilon}_2)(1 - \alpha)H(X|V)$.

Take $r > 0$ and let $D_n = \max_S \mathbb{D}(\tilde{P}_{WFZ_S} \| p_W^U p_F^U p_{Z_S})$ and $\mathcal{K}_n \triangleq \{D_n \geq r\}$. Using (43), we have that $\sum_{n=1}^{\infty} \mathbb{P}_{\mathcal{B}_n}(\mathcal{K}_n) < \infty$. Thus, using the first Borel-Cantelli lemma yields

$$\mathbb{P}_{\mathcal{B}_n}(\mathcal{K}_n \text{ infinitely often (i.o.)}) = 0. \quad (44)$$

This implies that, for all $r > 0$, $\mathbb{P}_{\mathcal{B}_n}(\{D_n < r\} \text{ i.o.}) = 1$, i.e., the sequence D_n converges to zero almost surely. Thus, the sequence D_n converges to zero in probability as well. We conclude that, for $R_s + \tilde{R}_s < (1 - \tilde{\epsilon}_2)(1 - \alpha)H(X|V)$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(\tilde{P}_{WFZ_S} \| p_W^U p_F^U p_{Z_S} \right) > 0 \right) = 0. \quad (45)$$

That is, protocol A is secure.

Next, we deduce that protocol B is also reliable and secure when $\tilde{R}_s \geq H(X|Y)$ and $R_s + \tilde{R}_s < (1 - \tilde{\epsilon}_2)(1 - \alpha)H(X|V)$. First, we show that the reliability in (36) holds for protocol B as well. We have

$$\begin{aligned} & \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}, P_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \\ & \leq \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S \hat{\mathbf{X}}} \right) \\ & \quad + \mathbb{V} \left(\tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}, P_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \\ & \leq \mathbb{V} \left(\tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \\ & \quad + \mathbb{V} \left(\tilde{P}_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\}, P_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \end{aligned} \quad (46)$$

$$\begin{aligned}
& + \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S \hat{\mathbf{X}}} \right) \\
& = \mathbb{V} \left(\tilde{P}_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \\
& \quad + 2\mathbb{V} \left(\tilde{P}_{WF}, P_W^U P_F^U \right), \tag{48}
\end{aligned}$$

where (46) and (47) follow from the triangle inequality, and (48) follows since (16), (17) and (25) imply that

$$\begin{aligned}
& \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}, \tilde{P}_{WFXYZ_S \hat{\mathbf{X}}} \right) \\
& = \mathbb{V} \left(\tilde{P}_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\}, P_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \tag{49} \\
& = \mathbb{V} \left(\tilde{P}_{WF}, P_W^U P_F^U \right). \tag{50}
\end{aligned}$$

Substituting (30) and (36) in (48) yields

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}, P_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) \right) = 0. \tag{51}$$

Second, we show that the secrecy property in (45) holds for protocol B. Using the union bound, we have

$$\begin{aligned}
& \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(P_{WFZ_S} \| P_W^U P_F^U P_{Z_S} \right) > 0 \right) \\
& \leq \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(P_{WFZ_S} \| P_W^U P_F^U P_{Z_S} \right) > 0, \right. \\
& \quad \left. \text{and } \mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) > 0 \right) \\
& \quad + \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(P_{WFZ_S} \| P_W^U P_F^U P_{Z_S} \right) > 0, \right. \\
& \quad \left. \text{and } \mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) = 0 \right) \tag{52} \\
& \leq \mathbb{P}_{\mathcal{B}_n} \left(\mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) > 0 \right) \\
& \quad + \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(\tilde{P}_{WFZ_S} \| P_W^U P_F^U P_{Z_S} \right) > 0 \right). \tag{53}
\end{aligned}$$

Equation (53) follows since $\mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) = 0$ if and only if $\tilde{P}_{WF}(w, f) = P_W^U P_F^U$ for all w and f , and hence $P_{WFZ_S} = P_W^U P_F^U P_{Z_S} |_{WF} = \tilde{P}_{WF} \tilde{P}_{Z_S} |_{WF} = \tilde{P}_{WFZ_S}$, where

$$\begin{aligned}
& P_{Z_S | WF}(\mathbf{z} | w, f) \\
& = \sum_{\mathbf{x} \in \mathcal{X}^n} P_{Z_S | \mathbf{X}}(\mathbf{z} | \mathbf{x}) \tilde{P}_{\mathbf{X} | WF}(\mathbf{x} | w, f) = \tilde{P}_{Z_S | WF}(\mathbf{z} | w, f). \tag{54}
\end{aligned}$$

Using the exponential decay in (30) and Markov inequality, we have, for any $r > 0$, that

$$\begin{aligned}
& \sum_{n=1}^{\infty} \mathbb{P}_{\mathcal{B}_n} \left(\mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) > r \right) \\
& \leq \frac{1}{r} \sum_{n=1}^{\infty} \mathbb{E}_{\mathcal{B}_n} \left(\mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) \right) \tag{55} \\
& \leq \frac{2}{r} \sum_{n=1}^{\infty} \exp(-\beta n) < \infty, \tag{56}
\end{aligned}$$

where $\beta > 0$. Thus, using the Borel-Cantelli lemma, as in the derivation for (45), we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{B}_n} \left(\mathbb{V}(\tilde{P}_{WF}, P_W^U P_F^U) > 0 \right) = 0. \tag{57}$$

By substituting (45) and (57) in (53), we get

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{B}_n} \left(\max_{S \in \mathcal{S}} \mathbb{D} \left(P_{WFZ_S} \| P_W^U P_F^U P_{Z_S} \right) > 0 \right) = 0. \tag{58}$$

Now, we show the existence of a binning realization, and hence an encoder and decoder, such that the reliability and secrecy properties, in (51) and (58), hold for protocol B. By applying Lemma 3 to the random sequence $\{\mathcal{B}_n\}_{n \geq 1}$ and the functions $\mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}, P_{WFXYZ_S} \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right)$, $\mathbb{1} \left\{ \max_{S \in \mathcal{S}} \mathbb{D} \left(P_{WFZ_S} \| P_W^U P_F^U P_{Z_S} \right) > 0 \right\}$, while using (51) and (58), there exists a sequence of binning realizations $\mathbf{b}_n^* = (b_{1,n}^*, b_{2,n}^*)$, with a corresponding joint distribution p^* for protocol B, such that

$$\lim_{n \rightarrow \infty} \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}^*, P_{WFXYZ_S}^* \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) = 0, \tag{59}$$

$$\lim_{n \rightarrow \infty} \mathbb{1} \left\{ \max_{S \in \mathcal{S}} \mathbb{D} \left(P_{WFZ_S}^* \| P_W^U P_F^U P_{Z_S} \right) > 0 \right\} = 0, \tag{60}$$

where $W = b_{1,n}^*(\mathbf{X}^n)$ and $F = b_{2,n}^*(\mathbf{X}^n)$.

Next, we introduce the random variable \hat{W} to the two joint distributions in (59), where \hat{W} is a deterministic function of the random sequence $\hat{\mathbf{X}}^n$, i.e., $p_{\hat{W} | \hat{\mathbf{X}}}^*(\hat{w} | \hat{\mathbf{x}}) = \mathbb{1} \left\{ \hat{w} = b_{1,n}^*(\hat{\mathbf{x}}) \right\}$. Using (25) and (59), we have

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}^*, P_{WFXYZ_S}^* \mathbb{1}\{\hat{W} = W\} \right) \\
& = \lim_{n \rightarrow \infty} \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}^* \mathbb{1}\{\hat{W} = b_{1,n}^*(\hat{\mathbf{X}})\}, \right. \\
& \quad \left. P_{WFXYZ_S}^* \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \mathbb{1}\{\hat{W} = b_{1,n}^*(\hat{\mathbf{X}})\} \right) \tag{61} \\
& = \lim_{n \rightarrow \infty} \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}^*, P_{WFXYZ_S}^* \mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\} \right) = 0, \tag{62}
\end{aligned}$$

where (61) follows since $p_{\hat{W} | WFXYZ_S \hat{\mathbf{X}}}^* = p_{\hat{W} | \hat{\mathbf{X}}}^* = \mathbb{1}\{\hat{W} = b_{1,n}^*(\hat{\mathbf{X}})\}$, and that $\hat{W} = W$ if and only if $\hat{\mathbf{X}} = \mathbf{X}$ and $\hat{W} = b_{1,n}^*(\hat{\mathbf{X}})$. We then have

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \mathbb{E}_F \left(\mathbb{P} p^*(\hat{W} \neq W | F) \right) \\
& = \lim_{n \rightarrow \infty} \sum_f P_F^U \sum_{w, \hat{w}: \hat{w} \neq w} P_{W \hat{W} | F}^*(w, \hat{w} | f) \tag{63} \\
& = \lim_{n \rightarrow \infty} \sum_{w, \hat{w}, f: \hat{w} \neq w} P_{W \hat{W} | F}^*(w, \hat{w}, f) \tag{64} \\
& = \lim_{n \rightarrow \infty} \sum_{\substack{w, \hat{w}, f: \\ P_W^U P_F^U \mathbb{1}\{\hat{w} = w\} \\ < p^*(w, \hat{w}, f)}} \left[P_{W \hat{W} | F}^*(w, \hat{w}, f) - P_W^U P_F^U \mathbb{1}\{\hat{w} = w\} \right] \tag{65} \\
& = \lim_{n \rightarrow \infty} \mathbb{V} \left(P_{W \hat{W} | F}^*, P_W^U P_F^U \mathbb{1}\{\hat{W} = W\} \right) \tag{66} \\
& = \lim_{n \rightarrow \infty} \mathbb{V} \left(P_{WFXYZ_S \hat{\mathbf{X}}}^*, P_{WFXYZ_S}^* \mathbb{1}\{\hat{W} = W\} \right) = 0. \tag{67}
\end{aligned}$$

Equation (65) follows because $p_{W \hat{W} | F}^* > P_W^U P_F^U \mathbb{1}\{\hat{W} = W\}$ if and only if $\mathbb{1}\{\hat{W} = W\} = 0$ where $p_{W \hat{W} | F}^*$ factorizes as $P_W^U P_F^U P_{\hat{W} | WF}^*$ and $p_{\hat{W} | WF}^* \leq 1$, while equation (67) follows from (25) and (62).

We also have that

$$\begin{aligned} & \mathbb{P}_F \left(\max_S \mathbb{D}(p_{WZ_S|F}^* \| p_W^U p_{Z_S|F}^*) > 0 \right) \\ & \leq \mathbb{P}_F \left(\max_S \mathbb{D}(p_{WZ_S|F}^* \| p_W^U p_{Z_S|F}^*) > 0, \right. \\ & \quad \text{and } \max_S \mathbb{D}(p_{WFZ_S}^* \| p_W^U p_F^U p_{Z_S}) = 0 \left. \right) \\ & + \mathbb{P}_F \left(\max_S \mathbb{D}(p_{WZ_S|F}^* \| p_W^U p_{Z_S|F}^*) > 0, \right. \\ & \quad \text{and } \max_S \mathbb{D}(p_{WFZ_S}^* \| p_W^U p_F^U p_{Z_S}) > 0 \left. \right) \quad (68) \end{aligned}$$

$$\begin{aligned} & \leq \mathbb{P}_F \left(\max_S \mathbb{D}(p_{WZ_S|F}^* \| p_W^U p_{Z_S|F}^*) > 0, \text{ and} \right. \\ & \quad \forall S, p_{WFZ_S}^*(w, f, \mathbf{z}) = p_W^U p_F^U p_{Z_S}(\mathbf{z}), \forall w, f, \mathbf{z} \left. \right) \\ & + \mathbb{P}_F \left(\max_S \mathbb{D}(p_{WFZ_S}^* \| p_W^U p_F^U p_{Z_S}) > 0 \right) \quad (69) \end{aligned}$$

$$= \mathbb{1} \left\{ \max_S \mathbb{D}(p_{WFZ_S}^* \| p_W^U p_F^U p_{Z_S}) > 0 \right\}, \quad (70)$$

where (69) follows since $\max_S \mathbb{D}(p_{WFZ_S}^* \| p_W^U p_F^U p_{Z_S}) = 0$, if and only if, for all $S \in \mathcal{S}$, $p_{WFZ_S}^*(w, f, \mathbf{z}) = p_W^U p_F^U p_{Z_S}(\mathbf{z})$ for all w, f , and \mathbf{z} . (70) follows because the first probability term on the right hand side of (69) is equal to zero. Thus, using (60), we get

$$\lim_{n \rightarrow \infty} \mathbb{P}_F \left(\max_{S \in \mathcal{S}} \mathbb{D}(p_{WZ_S|F}^* \| p_W^U p_{Z_S|F}^*) > 0 \right) = 0. \quad (71)$$

Let us express the random variable F as an explicit function of n , i.e., $F = F_n = b_{2,n}^*(\mathbf{X}^n)$. In order to eliminate F_n from the channel model in protocol B, we apply Lemma 3 to the random sequence $\{F_n\}_{n \geq 1}$ and the functions $\mathbb{P}_{p^*}(\hat{W} \neq W | F_n)$, $\mathbb{1} \left\{ \max_{S \in \mathcal{S}} \mathbb{D}(p_{WZ_S|F_n}^* \| p_W^U p_{Z_S|F_n}^*) > 0 \right\}$, while using (67) and (71), which implies that there exists at least one realization $\{f_n^*\}$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}_{p^*}(\hat{W} \neq W | F_n = f_n^*) = 0, \quad \text{and} \quad (72)$$

$$\lim_{n \rightarrow \infty} \max_{S \in \mathcal{S}} I_{p^*}(W; Z_S | F_n = f_n^*) = 0, \quad (73)$$

where I_{p^*} is the mutual information with respect to the distribution p^* . Equation (73) follows because $\lim_{n \rightarrow \infty} \mathbb{1} \left\{ \max_{S \in \mathcal{S}} \mathbb{D}(p_{WZ_S|F_n}^* \| p_W^U p_{Z_S|F_n}^*) > 0 \right\} = 0$ implies that there exists n' large enough such that, for all $n \geq n'$, we have

$$\begin{aligned} & \max_S \mathbb{D}(p_{WZ_S|F_n}^* \| p_W^U p_{Z_S|F_n}^*) \\ & = \max_S I_{p^*}(W; Z_S | F_n = f_n^*) = 0. \quad (74) \end{aligned}$$

Finally, let \tilde{p}^* be the induced distribution for protocol A corresponding to \mathbf{b}_n^* . We use $\tilde{p}_{\mathbf{X}|W, F_n=f_n^*}^*$ as the encoder and $(\tilde{p}_{\hat{\mathbf{X}}|\mathbf{Y}, F_n=f_n^*}^*, b_{1,n}^*(\hat{\mathbf{X}}))$ as the decoder for the original model. By combining the rate conditions $R_S + \tilde{R}_S < (1 - \tilde{\epsilon}_2)(1 - \alpha)H(X|V)$, $\tilde{R}_S \geq H(X|Y)$, and taking $\tilde{\epsilon}_2 \rightarrow 0$, the rate $R_S = \max_{p_X} [I(X; Y) - I(X; V) - \alpha H(X|V)]$ is achievable.

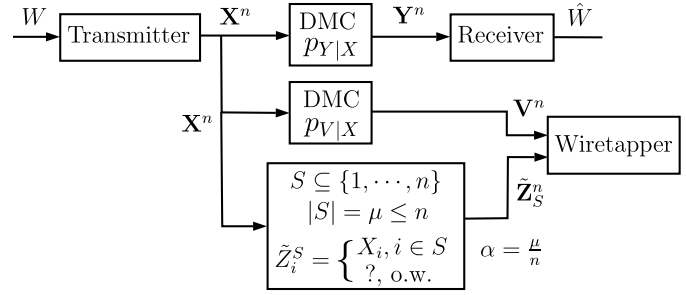


Fig. 3. A wiretap channel model whose secrecy capacity is equal to that of Figure 1.

So far, we have considered the case $U = X$. Next, we prefix a discrete memoryless channel $p_{X|U}$ to the original channel model in Figure 1. The main channel for the new model is $p_{Y|U}$ and the wiretapper channel is described by $p_{X|U}$ and (2). The proof for this case follows similar steps to the proof above. In particular, for protocol A, we consider the i.i.d. input sequence $\mathbf{U}^n = [U_1, U_2, \dots, U_n]$. When we apply Lemma 2 to protocol A, we set $\gamma = n(1 - \tilde{\epsilon}_2)[\alpha H(U|X) + (1 - \alpha)H(U|V)]$, and for $p_{U|Z_S}(\mathbf{u}|\mathbf{z}) > 0$, we have, for any $S \in \mathcal{S}$, that

$$p_{U|Z_S}(\mathbf{u}|\mathbf{z}) = p_{U_S U_{S^c} | \mathbf{X}_S \mathbf{V}_{S^c}}(\mathbf{u}_S, \mathbf{u}_{S^c} | \mathbf{x}_S, \mathbf{v}_{S^c}) \quad (75)$$

$$= p_{U_S | \mathbf{X}_S \mathbf{V}_{S^c}}(\mathbf{u}_S | \mathbf{x}_S, \mathbf{v}_{S^c}) \times p_{U_{S^c} | \mathbf{U}_S \mathbf{X}_S \mathbf{V}_{S^c}}(\mathbf{u}_{S^c} | \mathbf{u}_S, \mathbf{x}_S, \mathbf{v}_{S^c}) \quad (76)$$

$$= p_{U_S | \mathbf{X}_S}(\mathbf{u}_S | \mathbf{x}_S) p_{U_{S^c} | \mathbf{V}_{S^c}}(\mathbf{u}_{S^c} | \mathbf{v}_{S^c}) \quad (77)$$

$$= \prod_{i \in S} p_{U|X}(u_i | x_i) \prod_{i \in S^c} p_{U|V}(u_i | v_i), \quad (78)$$

where (77) and (78) follow since the sequences $\mathbf{U}^n, \mathbf{X}^n$, and \mathbf{V}^n are i.i.d. and the channels $p_{X|U}$ and $p_{V|X}$ are discrete memoryless channels. Using (78), the choice for γ , and Hoeffding's inequality, the conditions of Lemma 2 are satisfied, and we deduce the rate condition

$$R_S + \tilde{R}_S < (1 - \tilde{\epsilon}_2)[\alpha H(U|X) + (1 - \alpha)H(U|V)] \quad (79)$$

required for secrecy of protocol A. Note that $H(U|X) = H(U|X, V)$ because of the Markov chain $U - X - V$. By combining (79) with the rate condition $\tilde{R}_S \geq H(U|Y)$ required for the Slepian-Wolf decoder, we obtain the achievability of (7). The cardinality bound on \mathcal{U} , $|\mathcal{U}| \leq |\mathcal{X}|$, follows using [23, Appendix C]. This completes the achievability proof of Theorem 1.

V. CONVERSE

Consider the channel model illustrated in Figure 3, where the wiretapper observes the outputs of two independent channels, with \mathbf{X}^n being the input to both the channels. The first channel to the wiretapper is the DMC $p_{V|X}$ which outputs \mathbf{V}^n . The second channel is the wiretapper channel in the wiretap II channel model, i.e., the wiretapper chooses $S \subseteq [1, n]$ and observes $\tilde{\mathbf{Z}}_S^n = [\tilde{Z}_1^S, \dots, \tilde{Z}_n^S]$, where $\tilde{Z}_i^S = X_i$ for $i \in S$, and $\tilde{Z}_i^S = ?$, i.e., erasures, otherwise.

We show that, for $0 \leq \alpha \leq 1$, the strong secrecy capacity for this channel model, $C_S^{\text{EQ}}(\alpha)$, is equal to the strong secrecy

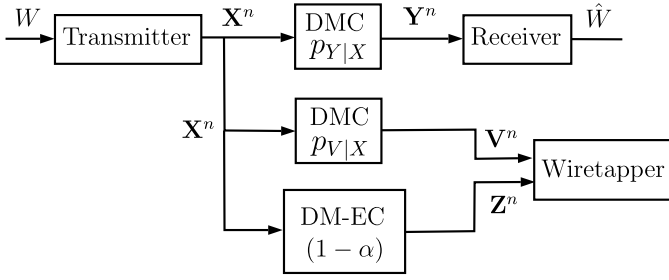


Fig. 4. A discrete memoryless equivalent wiretap channel model.

capacity of the original channel model, $C_s(\alpha)$, in (7). Since the main channels in the two models are the same, it suffices to show that $I(W; \mathbf{Z}_S^n) = I(W; \tilde{\mathbf{Z}}_S^n \mathbf{V}^n)$ for all $S \in \mathcal{S}$, where \mathbf{Z}_S^n is defined as in (2). This follows because, for all $S \in \mathcal{S}$, we have

$$H(W|\tilde{\mathbf{Z}}_S \mathbf{V})$$

$$= H(W, \mathbf{X}|\tilde{\mathbf{Z}}_S, \mathbf{V}) - H(\mathbf{X}|W, \tilde{\mathbf{Z}}_S, \mathbf{V}) \quad (80)$$

$$= H(\mathbf{X}|\tilde{\mathbf{Z}}_S, \mathbf{V}) + H(W|\mathbf{X}, \tilde{\mathbf{Z}}_S, \mathbf{V}) - H(\mathbf{X}|W, \tilde{\mathbf{Z}}_S, \mathbf{V}) \quad (81)$$

$$= H(\mathbf{X}|\tilde{\mathbf{Z}}_S, \mathbf{V}) - H(\mathbf{X}|W, \tilde{\mathbf{Z}}_S, \mathbf{V}) \quad (82)$$

$$= H(\mathbf{X}_S, \mathbf{X}_{S^c}|\mathbf{X}_S, \mathbf{V}_S, \mathbf{V}_{S^c}) - H(\mathbf{X}_S, \mathbf{X}_{S^c}|W, \mathbf{X}_S, \mathbf{V}_S, \mathbf{V}_{S^c}) \quad (83)$$

$$= H(\mathbf{X}_{S^c}|\mathbf{X}_S, \mathbf{V}_S, \mathbf{V}_{S^c}) - H(\mathbf{X}_{S^c}|W, \mathbf{X}_S, \mathbf{V}_S, \mathbf{V}_{S^c}) \quad (84)$$

$$= H(\mathbf{X}_{S^c}|\mathbf{X}_S, \mathbf{V}_{S^c}) - H(\mathbf{X}_{S^c}|W, \mathbf{X}_S, \mathbf{V}_{S^c}) \quad (85)$$

$$= H(\mathbf{X}|\mathbf{Z}_S) - H(\mathbf{X}|W, \mathbf{Z}_S) \quad (86)$$

$$= H(\mathbf{X}, W|\mathbf{Z}_S) - H(\mathbf{X}|W, \mathbf{Z}_S) = H(W|\mathbf{Z}_S), \quad (87)$$

where (82) and (87) follow because $H(W|\mathbf{X}) = 0$, and (85) follows since the channel $p_{V|X}$ is memoryless which results in the Markov chains $\mathbf{X}_{S^c} - \mathbf{X}_S \mathbf{V}_{S^c} - \mathbf{V}_S$ and $\mathbf{X}_{S^c} - W \mathbf{X}_S \mathbf{V}_{S^c} - \mathbf{V}_S$.

Next, consider the channel model illustrated in Figure 4, which is the same as the channel model in Figure 3, except we replace the second channel to the wiretapper with a discrete memoryless erasure channel (DM-EC) with erasure probability $1 - \alpha$. The output of the second channel to the wiretapper is \mathbf{Z}^n . For this model, we have the Markov chain $\mathbf{V}^n - \mathbf{X}^n - \mathbf{Z}^n$ since the two channels to the wiretapper are independent. Since the two channels to the wiretapper are discrete memoryless, we have

$$\begin{aligned} p_{\mathbf{V}\mathbf{Z}|\mathbf{X}}(\mathbf{v}, \mathbf{z}|\mathbf{x}) &= p_{\mathbf{V}|\mathbf{X}}(\mathbf{v}|\mathbf{x}) p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) \\ &= \prod_{i=1}^n p_{V|X}(v_i|x_i) p_{Z|X}(z_i|x_i) \\ &= \prod_{i=1}^n p_{VZ|X}(v_i, z_i|x_i). \end{aligned} \quad (88)$$

That is, the combined channel to the wiretapper is a discrete memoryless channel, making the channel model in Figure 4 a discrete memoryless wiretap channel. The strong secrecy capacity for this model $C_s^{\text{EQ2}}(\alpha)$ is given by

$$C_s^{\text{EQ2}}(\alpha) = \max_{U-X-YVZ} [I(U; Y) - I(U; VZ)]^+. \quad (89)$$

In order to compute $C_s^{\text{EQ2}}(\alpha)$ in (89), we define the random variable $\Phi \sim \text{Bern}(\alpha)$ whose n i.i.d. samples represent the

erasure process in the DM-EC, where $\Phi = 0$ when $Z = X$ and $\Phi = 1$ when $Z = ?$. Thus, Φ is determined by Z , and hence, the Markov chains $U - X - YVZ$ and $V - X - Z$ imply the Markov chains $U - X - YVZ\Phi$ and $V - X - Z\Phi$. Also, Φ is independent from X , since the erasure process is independent from the input to the channel. Thus, we have

$$\begin{aligned} p_{\Phi|UV}(\phi|u, v) &= \sum_{x \in \mathcal{X}} p_{\Phi X|UV}(\phi, x|u, v) \\ &= \sum_{x \in \mathcal{X}} p_{X|UV}(x|u, v) p_{\Phi|XUV}(\phi|x, u, v) \\ &= p_{\Phi}(\phi) \sum_{x \in \mathcal{X}} p_{X|UV}(x|u, v) = p_{\Phi}(\phi) \end{aligned} \quad (90)$$

$$\begin{aligned} p_{\Phi|V}(\phi|v) &= \sum_{x \in \mathcal{X}} p_{\Phi X|V}(\phi, x|v) \\ &= \sum_{x \in \mathcal{X}} p_{X|V}(x|v) p_{\Phi|XV}(\phi|x, v) \\ &= p_{\Phi}(\phi) \sum_{x \in \mathcal{X}} p_{X|V}(x|v) = p_{\Phi}(\phi). \end{aligned} \quad (91)$$

where (90) and (91) follow since $p_{\Phi|XUV} = p_{\Phi|XV} = p_{\Phi|X} = p_{\Phi}$ due to the Markov chains $U - XV - \Phi$ and $V - X - \Phi$, and the independence of Φ and X . Since $p_{\Phi|UV} = p_{\Phi|V} = p_{\Phi}$, then Φ and U are conditionally independent given V . Thus, we have

$$I(U; Z|V) = I(U; Z, \Phi|V) = I(U; Z|\Phi, V) \quad (92)$$

$$= \mathbb{P}(\Phi = 0)I(U; Z|\Phi = 0, V) + \mathbb{P}(\Phi = 1)I(U; Z|\Phi = 1, V) \quad (93)$$

$$= \alpha I(U; X|V) + (1 - \alpha)I(U; ?|V) \quad (94)$$

$$= \alpha I(U; X|V). \quad (95)$$

Substituting (95) in (89), we have

$$C_s^{\text{EQ2}}(\alpha) = \max_{U-X-YV} [I(U; Y) - I(U; V) - \alpha I(U; X|V)]^+. \quad (96)$$

Next, we use similar arguments to [9, Sec. V-C] to show that $C_s^{\text{EQ}}(\alpha) \leq C_s^{\text{EQ2}}(\alpha)$ for any $0 \leq \alpha \leq 1$ and sufficiently large n . The idea is that when the number of erasures of the DM-EC in the latter channel model (Figure 4) is greater than or equal to $(1 - \alpha)n$, the wiretapper's channel in the former (Figure 3) is better than her channel in the latter, since the wiretapper in the former is more capable and encounters a smaller number of erasures. Thus, $C_s^{\text{EQ}}(\alpha) \leq C_s^{\text{EQ2}}(\alpha)$ for this case. The result is established by using Sanov's theorem in method of types [24, Th. 11.4.1] to show that the probability that the DM-EC causes erasures less than $(1 - \alpha)n$ goes to 0 as $n \rightarrow \infty$.

In particular, we first show that, for $0 \leq \lambda < \alpha \leq 1$, we have $C_s^{\text{EQ}}(\alpha) \leq C_s^{\text{EQ2}}(\lambda)$. To do so, we show that every achievable strong secrecy rate for the channel model in Figure 3 is also achievable for the channel model in Figure 4. Fix λ such that $0 \leq \lambda < \alpha \leq 1$, and let R_s be an achievable strong secrecy rate for the former channel model. Thus, there exists a sequence

of $(n, 2^{nR_s})$ channel codes $\{\mathcal{C}_n^{\text{EQ}}\}_{n \geq 1}$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\hat{W} \neq W | \mathcal{C}_n^{\text{EQ}}) = 0, \quad (97)$$

$$\text{and } \lim_{n \rightarrow \infty} \max_{S \in \mathcal{S}} I(W; \tilde{\mathbf{Z}}_S, \mathbf{V} | \mathcal{C}_n^{\text{EQ}}) = 0. \quad (98)$$

We show that the rate R_s is also an achievable strong secrecy rate for the channel model in Figure 4 by showing that the sequence of $(n, 2^{nR_s})$ codes $\{\mathcal{C}_n^{\text{EQ}}\}_{n \geq 1}$ satisfies the constraints $\lim_{n \rightarrow \infty} \mathbb{P}(\hat{W} \neq W | \mathcal{C}_n^{\text{EQ}}) = 0$ and $\lim_{n \rightarrow \infty} I(W; \mathbf{Z}, \mathbf{V} | \mathcal{C}_n^{\text{EQ}}) = 0$ for this channel model.

The main channel in the two models is the same, and hence, the sequence of $(n, 2^{nR_s})$ codes $\{\mathcal{C}_n^{\text{EQ}}\}_{n \geq 1}$ achieves the reliability constraint for both channel models. Thus, it remains to show that $\{\mathcal{C}_n^{\text{EQ}}\}_{n \geq 1}$ achieves $\lim_{n \rightarrow \infty} I(W; \mathbf{Z}, \mathbf{V} | \mathcal{C}_n^{\text{EQ}}) = 0$.

Since $\lim_{n \rightarrow \infty} \max_{S \in \mathcal{S}} I(W; \tilde{\mathbf{Z}}_S, \mathbf{V}) = 0$, then for any $\epsilon_0 > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have

$$\begin{aligned} & \max_{S \in \mathcal{S}} I(W; \tilde{\mathbf{Z}}_S, \mathbf{V} | \mathcal{C}_n^{\text{EQ}}) \\ &= I(W; \mathbf{V} | \mathcal{C}_n^{\text{EQ}}) + \max_{S \in \mathcal{S}} I(W; \tilde{\mathbf{Z}}_S | \mathbf{V}, \mathcal{C}_n^{\text{EQ}}) \leq \frac{\epsilon_0}{2}. \end{aligned} \quad (99)$$

Let us define $\tilde{\mathcal{Z}} \triangleq \mathcal{X} \cup \{?\}$. For every $\mathbf{z}^n \in \tilde{\mathcal{Z}}^n$, define

$$\mathcal{N}(\mathbf{z}^n) \triangleq \{k \in \llbracket 1, n \rrbracket : z_k = ?\} \quad (100)$$

$$\Theta(\mathbf{z}^n) \triangleq \mathbb{1}\{|\mathcal{N}(\mathbf{z}^n)| \leq \lceil (1 - \alpha)n \rceil\}. \quad (101)$$

That is, $\mathcal{N}(\mathbf{z}^n)$ represents the number of erasures in the sequence \mathbf{z}^n , while $\Theta(\mathbf{z}^n)$ indicates whether the sequence \mathbf{z}^n has erasures less than or equal to $\lceil (1 - \alpha)n \rceil$.

For simplicity of notation, we drop $\mathcal{C}_n^{\text{EQ}}$ from the conditioning in (99); it is understood implicitly that the mutual information is calculated with respect to the code $\mathcal{C}_n^{\text{EQ}}$. Since $\Theta(\mathbf{Z}^n)$ is a deterministic function of \mathbf{Z}^n , the Markov chains $W - \mathbf{X}^n - \mathbf{V}^n \mathbf{Z}^n$ and $W \mathbf{V}^n - \mathbf{X}^n - \mathbf{Z}^n$ imply the Markov chains $W - \mathbf{X}^n - \mathbf{V}^n \mathbf{Z}^n \Theta(\mathbf{Z}^n)$ and $W \mathbf{V}^n - \mathbf{X}^n - \mathbf{Z}^n \Theta(\mathbf{Z}^n)$. Also, $\Theta(\mathbf{Z}^n)$ is independent from \mathbf{X}^n . Thus, we have

$$\begin{aligned} & p_{\Theta(\mathbf{Z}) | W \mathbf{V}}(\theta | w, \mathbf{v}) \\ &= \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\Theta(\mathbf{Z}) | \mathbf{X} W \mathbf{V}}(\theta, \mathbf{x} | w, \mathbf{v}) \\ &= \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\mathbf{X} | W \mathbf{V}}(\mathbf{x} | w, \mathbf{v}) p_{\Theta(\mathbf{Z}) | \mathbf{X} W \mathbf{V}}(\theta | \mathbf{x}, w, \mathbf{v}) \\ &= p_{\Theta(\mathbf{Z})}(\theta) \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\mathbf{X} | W \mathbf{V}}(\mathbf{x} | w, \mathbf{v}) = p_{\Theta(\mathbf{Z})}(\theta) \end{aligned} \quad (102)$$

$$\begin{aligned} & p_{\Theta(\mathbf{Z}) | \mathbf{V}}(\theta | \mathbf{v}) \\ &= \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\Theta(\mathbf{Z}) | \mathbf{X} \mathbf{V}}(\theta, \mathbf{x} | \mathbf{v}) \\ &= \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\mathbf{X} | \mathbf{V}}(\mathbf{x} | \mathbf{v}) p_{\Theta(\mathbf{Z}) | \mathbf{X} \mathbf{V}}(\theta | \mathbf{x}, \mathbf{v}) \\ &= p_{\Theta(\mathbf{Z})}(\theta) \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\mathbf{X} | \mathbf{V}}(\mathbf{x} | \mathbf{v}) = p_{\Theta(\mathbf{Z})}(\theta). \end{aligned} \quad (103)$$

From (102) and (103), W and $\Theta(\mathbf{Z})$ are conditionally

independent given \mathbf{V} , and hence,

$$I(W; \mathbf{Z} | \mathbf{V}) = I(W; \mathbf{Z}, \Theta(\mathbf{Z}) | \mathbf{V}) \quad (104)$$

$$= I(W; \Theta(\mathbf{Z}) | \mathbf{V}) + I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z})) \quad (105)$$

$$= I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z})) \quad (106)$$

$$= \mathbb{P}(\Theta(\mathbf{Z}) = 0) I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z}) = 0) \\ + \mathbb{P}(\Theta(\mathbf{Z}) = 1) I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z}) = 1). \quad (107)$$

We upper bound each term in the right hand side of (107). The first term is upper bounded by

$$\begin{aligned} & I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z}) = 0) \\ &= I(W; \mathbf{Z} | \mathbf{V}, \{|\mathcal{N}(\mathbf{Z})| > \lceil (1 - \alpha)n \rceil\}) \end{aligned} \quad (108)$$

$$\leq I(W; \mathbf{Z} | \mathbf{V}, \{|\mathcal{N}(\mathbf{Z})| = \lceil (1 - \alpha)n \rceil\}) \quad (109)$$

$$\leq \max_{S \in \mathcal{S}} I(W; \tilde{\mathbf{Z}}_S | \mathbf{V}). \quad (110)$$

We also have that

$$I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z}) = 1) \leq H(\mathbf{Z}) \leq n \log(|\mathcal{X}| + 1). \quad (111)$$

Next we upper bound $\mathbb{P}(\Theta(\mathbf{Z}) = 1)$. Take ν such that $\lambda < \nu < \alpha$, and hence, we have $\lceil (1 - \alpha)n \rceil \leq (1 - \nu)n < (1 - \lambda)n$. Let $\Phi_1, \Phi_2, \dots, \Phi_n$ be a sequence of i.i.d. binary random variables which represents the erasure process of the DM-EC in the model in Figure 4 ($\Phi_i = 1$ if $Z_i = X_i$, and $\Phi_i = 0$ if $Z_i = ?$), where Φ_i is distributed according to $Q_\Phi = \text{Bern}(\lambda)$. Let Q_Φ^n be the n -letter distribution of the sequence $\{\Phi_i\}_{i=1}^n$. For each $\zeta = \frac{l}{n}$, with $l \in \llbracket \lceil \nu n \rceil, n \rrbracket$, i.e., $\nu \leq \zeta < 1$, define the distribution $P_\Phi^{(\zeta)} = \text{Bern}(\zeta)$, and let \mathcal{P} be the set of all of these distributions. Let $T(P)$ denote the type class of the distribution P , i.e., all possible n -length sequences with the type (empirical distribution) P [24, Sec. 11.1]. Define the set $\mathcal{T} \triangleq \{T(P_\Phi^{(\zeta)}) : (1 - \zeta) \leq (1 - \nu)\}$. Using Sanov's theorem [24, Th. 11.4.1], we have

$$\mathbb{P}(\Theta(\mathbf{Z}) = 1) = \mathbb{P}_{Q_\Phi^n}(|\mathcal{N}(\mathbf{Z})| \leq \lceil (1 - \alpha)n \rceil) \quad (112)$$

$$\leq \mathbb{P}_{Q_\Phi^n}(|\mathcal{N}(\mathbf{Z})| \leq (1 - \nu)n) \quad (113)$$

$$= \mathbb{P}_{Q_\Phi^n}(\mathbb{1}\{k \in \llbracket 1, n \rrbracket : \Phi_k = 0\} \leq (1 - \nu)n) \quad (114)$$

$$= \mathbb{P}_{Q_\Phi^n}(\mathcal{T}) = \mathbb{P}_{Q_\Phi^n}(\mathcal{P}) \leq (n + 1)^2 2^{-n \mathbb{D}(P_\Phi^* || Q_\Phi)}, \quad (115)$$

where

$$\begin{aligned} & P_\Phi^* = \underset{P_\Phi^{(\zeta)} \in \mathcal{P}}{\text{argmin}} \mathbb{D}(P_\Phi^{(\zeta)} || Q_\Phi) \\ &= \underset{\zeta: \zeta \geq \nu}{\text{argmin}} \left(\zeta \log \frac{\zeta}{\lambda} + (1 - \zeta) \frac{1 - \zeta}{1 - \lambda} \right) = \text{Bern}(\nu). \end{aligned} \quad (116)$$

Note that $\mathbb{D}(P_\Phi^* || Q_\Phi) > 0$ since $\nu \neq \lambda$.

Using (111) and (115), the second term in the right hand side of (107) is upper bounded by

$$\log(|\mathcal{X}| + 1) n (n + 1)^2 2^{-n \mathbb{D}(P_\Phi^* || Q_\Phi)} \xrightarrow{n \rightarrow \infty} 0. \quad (117)$$

Thus, for $\epsilon_0 > 0$, there exists $n_1 \in \mathbb{N}$ such that, for all $n \geq n_1$,

$$\mathbb{P}(\Theta(\mathbf{Z}) = 1) I(W; \mathbf{Z} | \mathbf{V}, \Theta(\mathbf{Z}) = 1) \leq \frac{\epsilon_0}{2}. \quad (118)$$

Using (99), (107), (110), and (118), we have, for sufficiently large n , that

$$I(W; \mathbf{Z}, \mathbf{V}|c_n^{\text{EQ}}) = I(W; \mathbf{V}|c_n^{\text{EQ}}) + I(W; \mathbf{Z}|\mathbf{V}, c_n^{\text{EQ}}) \quad (119)$$

$$\leq I(W; \mathbf{V}|c_n^{\text{EQ}}) + \max_{S \in \mathcal{S}} I(W; \tilde{\mathbf{Z}}_S|\mathbf{V}, c_n^{\text{EQ}}) + \frac{\epsilon_0}{2} \leq \epsilon_0. \quad (120)$$

Thus, for $0 \leq \lambda < \alpha \leq 1$, we have $C_s^{\text{EQ}}(\alpha) \leq C_s^{\text{EQ}2}(\lambda)$. The right hand side of (96) is a continuous function of α , for $0 < \alpha < 1$ [9, Lemma 6]. Thus, by taking $\lambda \rightarrow \alpha$, we have $C_s^{\text{EQ}}(\alpha) \leq C_s^{\text{EQ}2}(\alpha)$. Note that for $\alpha = 0, 1$, we have $C_s^{\text{EQ}}(\alpha) = C_s^{\text{EQ}2}(\alpha)$. Thus, the secrecy capacity of the original model in Figure 1 is upper bounded by (96). This completes the proof for Theorem 1.

VI. DISCUSSION

In the converse proof for Theorem 1, we have shown that the strong secrecy capacity $C_s(\alpha)$ for the new wiretap channel model is equal to the strong secrecy capacity when the wiretapper, in addition to observing μ transmitted symbols of her choice noiselessly, observes the whole sequence \mathbf{V}^n . This is not surprising since observing noisy versions of the transmitted symbols through the DMC $p_{V|X}$ in the same positions where noiseless versions are available does not increase the wiretapper's information about the message. The expression for the strong secrecy capacity in (7) is thus intuitive where $I(U, V)$ represents the secrecy cost due to observing the whole sequence \mathbf{V}^n , and $\alpha I(U; X|V)$ represents the secrecy cost due to observing a fraction α of the transmitted symbols noiselessly, given the wiretapper's knowledge of the V outputs in these positions. Furthermore, the alternative characterization for the strong secrecy capacity in (8) is again intuitively pleasing, where the overall secrecy cost is represented by a weighted sum of the secrecy costs due to the noiseless and the noisy observations at the wiretapper, i.e., $\alpha I(U; X)$ and $(1 - \alpha)I(U; V)$.

It is worth noting that a problem similar to the model considered in this paper appears in the context of Quantum Cryptography when a transmitter and a receiver wish to agree on a secret key over a quantum channel in the presence of an external adversary [25], [26]. The adversary can apply any arbitrary sequence of operations, allowed by the laws of quantum physics, on the quantum states exchanged between the transmitter and receiver. The security of the key follows from the impossibility of applying such operations on a quantum mechanical system without changing its overall state. The legitimate terminals, by communicating over an additional classical error-free channel, can estimate the number of errors in the system, caused by the adversary, and abort the key agreement protocol if the number of errors exceeds a certain threshold. To sum up, like the models considered in this paper, the adversary in the problem described above is limited only in the fraction of time of being active. We refer the reader to [27], [28], and references therein, for a comprehensive treatment of the problems and utilized tools in quantum information theory.

Finally, we note that extending the proposed achievability approach in this paper to the case of a non-uniform message at the transmitter, i.e., semantic secrecy, does not appear straightforward. In particular, in order to handle the case of a non-uniform message, we would need to characterize the distribution of the source \mathbf{X}^n given the wiretapper's observation \mathbf{Z}_S^n , when conditioned over each key realization w , i.e., $p_{\mathbf{X}^n|\mathbf{Z}_S^n, w}(\mathbf{x}|\mathbf{z}, w)$, for all $\mathbf{x} \in \mathcal{X}^n, \mathbf{z} \in \mathcal{Z}^n$, and $w \in \llbracket 1, 2^{nR_s} \rrbracket$, which is not easy.

VII. CONCLUSION

In this work, we have introduced a wiretap channel model that generalizes the classical wiretap channel models, and derived its strong secrecy capacity. The model generalizes the classical wiretap channel [2], [3] to one with a wiretapper who chooses a fixed-length subset of the transmitted codeword symbols to perfectly access, and generalizes the wiretap channel II with a discrete memoryless main channel in [8] to one with a wiretapper who observes the output of a noisy channel instead of the erasures. The wiretapper in this model is still passive, yet she is more capable than a classical wiretapper since she can tap a subset of the symbols of her choosing noiselessly, while still receiving the remaining symbols through a channel. Our secrecy capacity result quantifies the secrecy cost of this additional capability of the wiretapper, with respect to the previous wiretap channel models.

As for current and future directions, exploring the multi-terminal extensions of this new model is the natural next step, similar to multi-terminal extensions for Wyner's original model, e.g., [29]–[34]. Additionally, it is of interest to seek new and more powerful wiretapper models against which information theoretic security guarantees can be established.

APPENDIX A PROOF OF LEMMA 1

For $w, f \in \llbracket 1, \tilde{W} \rrbracket \times \llbracket 1, \tilde{F} \rrbracket$, we have

$$P_{WF}(w, f) = \sum_{x \in \mathcal{X}} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}. \quad (121)$$

We also have that, for all $x \in \mathcal{X}$,

$$\begin{aligned} \mathbb{E}_{\mathcal{B}} \left(\mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \right) \\ = \mathbb{P}(\mathcal{B}_1(x) = w) \mathbb{P}(\mathcal{B}_2(x) = f) = \frac{1}{\tilde{W}\tilde{F}}. \end{aligned} \quad (122)$$

Thus, we have $\mathbb{E}_{\mathcal{B}}(P_{WF}) = \frac{1}{\tilde{W}\tilde{F}} = p_W^U p_F^U$. For all w and f , define the random variables

$$P_1(w, f) = \sum_{x \notin \mathcal{D}_\gamma} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \quad (123)$$

$$P_2(w, f) = \sum_{x \in \mathcal{D}_\gamma} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}. \quad (124)$$

Note that $P_{WF}(w, f) = P_1(w, f) + P_2(w, f)$. Thus, we have

$$\begin{aligned} & \mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(P_{WF}, p_{WF}^U \right) \right) \\ &= \frac{1}{2} \mathbb{E}_{\mathcal{B}} \left(\sum_{w,f} \left| P_{WF}(w, f) - \mathbb{E}_{\mathcal{B}}(P_{WF}(w, f)) \right| \right) \end{aligned} \quad (125)$$

$$= \frac{1}{2} \mathbb{E}_{\mathcal{B}} \left(\sum_{w,f} \left| \sum_{i=1}^2 (P_i(w, f) - \mathbb{E}_{\mathcal{B}}(P_i(w, f))) \right| \right) \quad (126)$$

$$\begin{aligned} & \leq \frac{1}{2} \sum_{w,f} \mathbb{E}_{\mathcal{B}} |P_1(w, f) - \mathbb{E}_{\mathcal{B}}(P_1(w, f))| \\ & \quad + \frac{1}{2} \sum_{w,f} \mathbb{E}_{\mathcal{B}} |P_2(w, f) - \mathbb{E}_{\mathcal{B}}(P_2(w, f))|, \end{aligned} \quad (127)$$

where (127) follows from the triangle inequality. We now upper bound each term on the right hand side of (127). For the first term, we have

$$\begin{aligned} & \frac{1}{2} \sum_{w,f} \mathbb{E}_{\mathcal{B}} \left| P_1(w, f) - \mathbb{E}_{\mathcal{B}}(P_1(w, f)) \right| \\ & \leq \sum_{w,f} \mathbb{E}_{\mathcal{B}}(P_1(w, f)) \end{aligned} \quad (128)$$

$$= \sum_{w,f} \sum_{x \notin \mathcal{D}_\gamma} p_X(x) \mathbb{E}_{\mathcal{B}} \left(\mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \right) \quad (129)$$

$$= \sum_{x \notin \mathcal{D}_\gamma} p_X(x) = \mathbb{P}(X \notin \mathcal{D}_\gamma), \quad (130)$$

where (128) follows from the triangle inequality.

For the second term in the right hand side of (127), we have

$$\begin{aligned} & \frac{1}{2} \sum_{w,f} \mathbb{E}_{\mathcal{B}} \left| P_2(w, f) - \mathbb{E}_{\mathcal{B}}(P_2(w, f)) \right| \\ & = \frac{1}{2} \sum_{w,f} \mathbb{E}_{\mathcal{B}} \sqrt{(P_2(w, f) - \mathbb{E}_{\mathcal{B}}(P_2(w, f)))^2} \end{aligned} \quad (131)$$

$$\leq \frac{1}{2} \sum_{w,f} \sqrt{\mathbb{E}_{\mathcal{B}}(P_2(w, f) - \mathbb{E}_{\mathcal{B}}(P_2(w, f)))^2} \quad (132)$$

$$= \frac{1}{2} \sum_{w,f} \sqrt{\text{Var}_{\mathcal{B}}(P_2(w, f))} \leq \frac{1}{2} \sqrt{\frac{\tilde{W}\tilde{F}}{2^\gamma}}, \quad (133)$$

where (132) follows from Jensen's inequality and the concavity of square root. The inequality in (133) follows because, for all w and f , we have

$$\begin{aligned} & \text{Var}_{\mathcal{B}}(P_2(w, f)) \\ & = \text{Var}_{\mathcal{B}} \left(\sum_{x \in \mathcal{D}_\gamma} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \right) \end{aligned} \quad (134)$$

$$= \sum_{x \in \mathcal{D}_\gamma} \text{Var}_{\mathcal{B}} \left(p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \right) \quad (135)$$

$$\leq \sum_{x \in \mathcal{D}_\gamma} p_X^2(x) \mathbb{E}_{\mathcal{B}} \left(\mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \right) \quad (136)$$

$$= \frac{1}{\tilde{W}\tilde{F}} \sum_{x \in \mathcal{D}_\gamma} p_X^2(x) \quad (137)$$

$$\leq \frac{2^{-\gamma}}{\tilde{W}\tilde{F}} \sum_{x \in \mathcal{D}_\gamma} p_X(x) \leq \frac{2^{-\gamma}}{\tilde{W}\tilde{F}}, \quad (138)$$

where (135) follows since the random variables $\{p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}\}_{x \in \mathcal{D}_\gamma}$ are independent due to the structure of the random binning, and (138) follows because $p_X(x) \leq 2^{-\gamma}$ for all $x \in \mathcal{D}_\gamma$. Lemma 1 follows from substituting (130) and (133) in (127).

APPENDIX B PROOF OF LEMMA 2

We first state the following lemma, which provides a variation of Chernoff bound.

Lemma 6: (A variation on Chernoff bound:) Let U_1, U_2, \dots, U_n be a sequence of non-negative independent random variables with respective means $\mathbb{E}(U_i) = \bar{m}_i$. If $U_i \in [0, b]$, for all $i \in \llbracket 1, n \rrbracket$, and $\sum_{i=1}^n \bar{m}_i \leq \bar{m}$, then, for every $\epsilon \in [0, 1]$, we have

$$\mathbb{P} \left(\sum_{i=1}^n U_i \geq (1 + \epsilon) \bar{m} \right) \leq \exp \left(-\epsilon^2 \frac{\bar{m}}{3b} \right). \quad (139)$$

Proof: The proof is adapted from [9, Appendix C]. The details are relegated to Appendix C. ■

A. High Probability \mathcal{Z} -Set

For all $S \in \mathcal{S}$, define the set

$$\mathcal{A}_S \triangleq \left\{ z \in \mathcal{Z} : \mathbb{P}_{p_{X|Z_S}} \left((X, z) \in \mathcal{D}_\gamma^S \right) \geq 1 - \delta \right\}. \quad (140)$$

Recall that $\mathbb{P}_{p_{XZ_S}} \left((X, Z_S) \in \mathcal{D}_\gamma^S \right) \geq 1 - \delta^2$ by assumption. Using Markov inequality, we have

$$\mathbb{P}_{p_{Z_S}}(\mathcal{A}_S^c) = \mathbb{P}_{p_{Z_S}} \left(\mathbb{P}_{p_{X|Z_S}} \left((X, Z_S) \notin \mathcal{D}_\gamma^S \right) \geq \delta \right) \quad (141)$$

$$\leq \frac{1}{\delta} \mathbb{E}_{p_{Z_S}} \left(\mathbb{P}_{p_{X|Z_S}} \left((X, Z_S) \notin \mathcal{D}_\gamma^S \right) \right) \quad (142)$$

$$= \frac{1}{\delta} \mathbb{P}_{p_{XZ_S}} \left((X, Z_S) \notin \mathcal{D}_\gamma^S \right) \quad (143)$$

$$\leq \frac{\delta^2}{\delta} = \delta. \quad (144)$$

B. Typical and Non-Typical Events

For all $w, f \in \llbracket 1, \tilde{W} \rrbracket \times \llbracket 1, \tilde{F} \rrbracket$, $z \in \mathcal{Z}$, and $S \in \mathcal{S}$, define the random variables

$$\begin{aligned} P_1^S(w, f|z) &= \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\{\mathcal{B}_1(x) = w\} \\ & \quad \times \mathbb{1}\{\mathcal{B}_2(x) = f\} \mathbb{1} \left\{ (x, z) \in \mathcal{D}_\gamma^S \right\} \end{aligned} \quad (145)$$

$$\begin{aligned} P_2^S(w, f|z) &= \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\{\mathcal{B}_1(x) = w\} \\ & \quad \times \mathbb{1}\{\mathcal{B}_2(x) = f\} \mathbb{1} \left\{ (x, z) \notin \mathcal{D}_\gamma^S \right\}. \end{aligned} \quad (146)$$

Thus, we have, for all w, f, z , and S , that

$$\begin{aligned} P_{WF|Z_S}(w, f|z) &= \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \quad (147) \\ &= P_1^S(w, f|z) + P_2^S(w, f|z). \quad (148) \end{aligned}$$

Note that, for fixed $z \in \mathcal{Z}$ and $S \in \mathcal{S}$, each of the the random variables $P_i^S(w, f|z)$, $i = 1, 2$, is identically distributed for all $w, f \in \llbracket 1, \tilde{W} \rrbracket \times \llbracket 1, \tilde{F} \rrbracket$ due to the symmetry in the random binning. We then fix $z \in \mathcal{Z}$ and $S \in \mathcal{S}$, and let $P_1^S(w, f|z) = \sum_{x \in \mathcal{X}} U_x(w, f, z, S)$, where

$$\begin{aligned} U_x(w, f, z, S) &= p_{X|Z_S}(x|z) \mathbb{1}\{\mathcal{B}_1(x) = w\} \\ &\quad \times \mathbb{1}\{\mathcal{B}_2(x) = f\} \mathbb{1}\{(x, z) \in \mathcal{D}_\gamma^S\}. \quad (149) \end{aligned}$$

The random variables $\{U_x(w, f, z, S)\}_{x \in \mathcal{X}}$ are non-negative and independent, and for all $x \in \mathcal{X}$,

$$U_x(w, f, z, S) \leq p_{X|Z_S}(x|z) \mathbb{1}\{(x, z) \in \mathcal{D}_\gamma^S\} < 2^{-\gamma}, \quad (150)$$

where $p_{X|Z_S}(x|z) < 2^{-\gamma}$, for all $(x, z) \in \mathcal{D}_\gamma^S$. Also, we have

$$\begin{aligned} \sum_{x \in \mathcal{X}} \mathbb{E}_{\mathcal{B}}(U_x(w, f, z, S)) &= \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{E}_{\mathcal{B}}(\mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}) \\ &\quad \times \mathbb{1}\{(x, z) \in \mathcal{D}_\gamma^S\} \quad (151) \end{aligned}$$

$$= \frac{1}{\tilde{W}\tilde{F}} \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\{(x, z) \in \mathcal{D}_\gamma^S\} \quad (152)$$

$$= \frac{1}{\tilde{W}\tilde{F}} \mathbb{P}_{p_{X|Z_S}}((X, z) \in \mathcal{D}_\gamma^S). \quad (153)$$

By applying Lemma 6 to the random variables $\{U_x(w, f, z, S)\}_{x \in \mathcal{X}}$, with $\bar{m} = \frac{\mathbb{P}_{p_{X|Z_S}}((X, z) \in \mathcal{D}_\gamma^S)}{\tilde{W}\tilde{F}}$ and $b = 2^{-\gamma}$, we have, for every $\epsilon_1 \in [0, 1]$ and $z \in \mathcal{A}_S$, that

$$\begin{aligned} \mathbb{P}_{\mathcal{B}}\left(P_1^S(w, f|z) \geq \frac{1 + \epsilon_1}{\tilde{W}\tilde{F}}\right) &\leq \mathbb{P}\left(\sum_{x \in \mathcal{X}} U_x(w, f, z, S) \geq \frac{1 + \epsilon_1}{\tilde{W}\tilde{F}} \mathbb{P}_{p_{X|Z_S}}((X, z) \in \mathcal{D}_\gamma^S)\right) \quad (154) \end{aligned}$$

$$\leq \exp\left(\frac{-\epsilon_1^2 \mathbb{P}_{p_{X|Z_S}}((X, z) \in \mathcal{D}_\gamma^S) 2^\gamma}{3\tilde{W}\tilde{F}}\right) \quad (155)$$

$$\leq \exp\left(\frac{-\epsilon_1^2(1 - \delta)2^\gamma}{3\tilde{W}\tilde{F}}\right), \quad (156)$$

where (154) follows since $\mathbb{P}_{p_{X|Z_S}}((X, z) \in \mathcal{D}_\gamma^S) \leq 1$, and (156) follows because, for all $z \in \mathcal{A}_S$, we have $\mathbb{P}_{p_{X|Z_S}}((X, z) \in \mathcal{D}_\gamma^S) \geq 1 - \delta$.

We also have have that,

$$\begin{aligned} \mathbb{E}_{p_{Z_S}}\left(\sum_{w, f} P_2^S(w, f|Z_S)\right) &= \mathbb{E}_{p_{Z_S}}\left(\sum_{x \in \mathcal{X}} p_{X|Z_S}(x|Z_S) \mathbb{1}\{(x, Z_S) \notin \mathcal{D}_\gamma^S\}\right. \\ &\quad \left. \times \sum_{w, f} \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}\right) \quad (157) \end{aligned}$$

$$= \sum_{z \in \mathcal{Z}} p_{Z_S}(z) \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\{(x, z) \notin \mathcal{D}_\gamma^S\} \quad (158)$$

$$= \sum_{(x, z) \notin \mathcal{D}_\gamma^S} p_{XZ_S}(x, z) \quad (159)$$

$$= \mathbb{P}_{p_{XZ_S}}((X, Z_S) \notin \mathcal{D}_\gamma^S) \leq \delta^2, \quad (160)$$

where (158) follows since every $x \in \mathcal{X}$ is assigned to only one pair (w, f) , and hence,

$$\sum_{w, f} \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} = 1. \quad (161)$$

C. Good Binning Functions

Let $\mathbf{b} \triangleq (b_1, b_2) : \mathcal{X} \mapsto \llbracket 1, \tilde{W} \rrbracket \times \llbracket 1, \tilde{F} \rrbracket$ be a realization of the random binning \mathcal{B} . Recall that the random variable $P_1^S(w, f|z)$ is identically distributed for every w and f . We then define the class \mathcal{G} of binning functions \mathbf{b} as

$$\mathcal{G} \triangleq \left\{ \mathbf{b} : P_1^S(w, f|z) < \frac{1 + \epsilon_1}{\tilde{W}\tilde{F}}, \right. \\ \left. \text{for all } S \in \mathcal{S} \text{ and } z \in \mathcal{A}_S \right\}. \quad (162)$$

Using the union bound and (156), we have that

$$\begin{aligned} \mathbb{P}_{\mathcal{B}}(\mathcal{G}^c) &= \mathbb{P}_{\mathcal{B}}\left(P_1^S(w, f|z) \geq \frac{1 + \epsilon_1}{\tilde{W}\tilde{F}}, \right. \\ &\quad \left. \text{for some } S \in \mathcal{S}, \text{ or } z \in \mathcal{A}_S\right) \quad (163) \end{aligned}$$

$$= \mathbb{P}_{\mathcal{B}}\left(\bigcup_{S \in \mathcal{S}} \bigcup_{z \in \mathcal{A}_S} P_1^S(w, f|z) \geq \frac{1 + \epsilon_1}{\tilde{W}\tilde{F}}\right) \quad (164)$$

$$\leq \sum_{S \in \mathcal{S}} \sum_{z \in \mathcal{A}_S} \mathbb{P}_{\mathcal{B}}\left(P_1^S(w, f|z) \geq \frac{1 + \epsilon_1}{\tilde{W}\tilde{F}}\right) \quad (165)$$

$$\leq \sum_{S \in \mathcal{S}} |\mathcal{A}_S| \exp\left(\frac{-\epsilon_1^2(1 - \delta)2^\gamma}{3\tilde{W}\tilde{F}}\right) \quad (166)$$

$$\leq |\mathcal{S}| |\mathcal{Z}| \exp\left(\frac{-\epsilon_1^2(1 - \delta)2^\gamma}{3\tilde{W}\tilde{F}}\right). \quad (167)$$

Take \mathbf{b} such that $\mathbf{b} \in \mathcal{G}$, and set $W = b_1(X)$ and $F = b_2(X)$. For every $S \in \mathcal{S}$, we have

$$\begin{aligned} & \mathbb{D}\left(P_{WFZ_S} \parallel P_W^U P_F^U P_{Z_S}\right) \\ &= \mathbb{E}_{P_{Z_S}}\left(\mathbb{D}\left(P_{WF|Z_S} \parallel P_W^U P_F^U\right)\right) \end{aligned} \quad (168)$$

$$= \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_{WF|Z_S}(w, f|Z_S) \log \frac{P_{WF|Z_S}(w, f|Z_S)}{P_W^U P_F^U}\right) \quad (169)$$

$$= \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} \sum_{i=1}^2 P_i^S(w, f|Z_S) \times \log\left(\tilde{W}\tilde{F} \sum_{i=1}^2 P_i^S(w, f|Z_S)\right)\right) \quad (170)$$

$$= \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} \sum_{i=1}^2 P_i^S(w, f|Z_S) \times \log \frac{\sum_{i=1}^2 P_i^S(w, f|Z_S)}{\frac{1}{\tilde{W}\tilde{F}} \sum_{i=1}^2 \sum_{w,f} P_i^S(w, f|Z_S)}\right) \quad (171)$$

$$\leq \mathbb{E}_{P_{Z_S}}\left(\sum_{i=1}^2 \sum_{w,f} P_i^S(w, f|Z_S) \log \frac{\tilde{W}\tilde{F} P_i^S(w, f|Z_S)}{\sum_{w,f} P_i^S(w, f|Z_S)}\right) \quad (172)$$

$$\begin{aligned} &= \sum_{i=1}^2 \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_i^S(w, f|Z_S) \log\left(\tilde{W}\tilde{F} P_i^S(w, f|Z_S)\right)\right) \\ &+ \mathbb{E}_{P_{Z_S}}\left(\sum_{i=1}^2 \sum_{w,f} P_i^S(w, f|Z_S) \log \frac{1}{\sum_{w,f} P_i^S(w, f|Z_S)}\right) \end{aligned} \quad (173)$$

where (171) follows because

$$\sum_{w,f} \sum_{i=1}^2 P_i^S(w, f|Z_S) = \sum_{w,f} P_{WF|Z_S}(w, f|Z_S) = 1, \quad (174)$$

and (172) follows from the log-sum inequality.

Now, we upper bound each term in (173). For $\mathbf{b} \in \mathcal{G}$ and every $S \in \mathcal{S}$, we have

$$\begin{aligned} & \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_1^S(w, f|Z_S) \log\left(\tilde{W}\tilde{F} P_1^S(w, f|Z_S)\right)\right) \\ &= \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_1^S(w, f|Z_S) \times \log\left(\tilde{W}\tilde{F} P_1^S(w, f|Z_S)\right) \mathbb{1}\{Z_S \in \mathcal{A}_S\}\right) \\ &+ \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_1^S(w, f|Z_S) \times \log\left(\tilde{W}\tilde{F} P_1^S(w, f|Z_S)\right) \mathbb{1}\{Z_S \notin \mathcal{A}_S\}\right) \end{aligned} \quad (175)$$

$$\begin{aligned} & < \log(1 + \epsilon_1) + \sum_{x,z} p_{XZ_S}(x, z) \\ & \times \log\left(\tilde{W}\tilde{F} P_1^S(w, f|z)\right) \mathbb{1}\{z \notin \mathcal{A}_S\} \end{aligned} \quad (176)$$

$$\leq \log(1 + \epsilon_1) + \log(\tilde{W}\tilde{F}) \mathbb{P}_{P_{Z_S}}(Z_S \notin \mathcal{A}_S) \quad (177)$$

$$\leq \epsilon_1 + \delta \log(\tilde{W}\tilde{F}), \quad (178)$$

where (176) follows because, for every $\mathbf{b} \in \mathcal{G}$ and $S \in \mathcal{S}$, we have $\tilde{W}\tilde{F} P_1^S(w, f|Z_S) < (1 + \epsilon)$ for $Z_S \in \mathcal{A}_S$ and every w, f , and (178) follows from (144).

Using (160), we have, for every $S \in \mathcal{S}$, that

$$\begin{aligned} & \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_2^S(w, f|Z_S) \log\left(\tilde{W}\tilde{F} P_2^S(w, f|Z_S)\right)\right) \\ & \leq \log(\tilde{W}\tilde{F}) \mathbb{E}_{P_{Z_S}}\left(\sum_{w,f} P_2^S(w, f|Z_S)\right) \leq \delta^2 \log(\tilde{W}\tilde{F}). \end{aligned} \quad (179)$$

We also have, for every $S \in \mathcal{S}$, that

$$\begin{aligned} & \mathbb{E}_{P_{Z_S}}\left(\sum_{i=1}^2 \sum_{w,f} P_i^S(w, f|Z_S) \log \frac{1}{\sum_{w,f} P_i^S(w, f|Z_S)}\right) \\ &= \mathbb{E}_{P_{Z_S}}\left(H_b\left(\mathbb{P}_{P_{X|Z_S}}\left((X, Z_S) \in \mathcal{D}_\gamma^S\right)\right)\right) \end{aligned} \quad (180)$$

$$\leq H_b\left(\mathbb{E}_{P_{Z_S}}\left(\mathbb{P}_{P_{X|Z_S}}\left((X, Z_S) \in \mathcal{D}_\gamma^S\right)\right)\right) \quad (181)$$

$$= H_b\left(\mathbb{P}_{P_{XZ_S}}\left((X, Z_S) \in \mathcal{D}_\gamma^S\right)\right) \quad (182)$$

$$\leq H_b(1 - \delta^2) = H_b(\delta^2), \quad (183)$$

where (181) follows from Jensen's inequality and the concavity of H_b , and (183) follows since $H_b(x)$ is monotonically decreasing in $x \in (\frac{1}{2}, 1)$. Equation (180) follows since $\sum_{i=1}^2 \sum_{w,f} P_i^S(w, f|Z_S) = 1$, and $\sum_{w,f} P_1^S(w, f|Z_S) = \mathbb{P}_{P_{X|Z_S}}\left((X, Z_S) \in \mathcal{D}_\gamma^S\right)$.

By substituting (178), (179), and (183) in (173), we have, for every $\mathbf{b} \in \mathcal{G}$ and $S \in \mathcal{S}$, that

$$\begin{aligned} & \mathbb{D}\left(P_{WFZ_S} \parallel P_W^U P_F^U P_{Z_S}\right) \\ & < \epsilon_1 + (\delta + \delta^2) \log(\tilde{W}\tilde{F}) + H_b(\delta^2) = \tilde{\epsilon}. \end{aligned} \quad (184)$$

Thus, we have

$$\begin{aligned} & \mathbb{P}_{\mathcal{B}}\left(\max_{S \in \mathcal{S}} \mathbb{D}\left(P_{WFZ_S} \parallel P_W^U P_F^U P_{Z_S}\right) \geq \tilde{\epsilon}\right) \\ &= 1 - \mathbb{P}_{\mathcal{B}}\left(\max_{S \in \mathcal{S}} \mathbb{D}\left(P_{WFZ_S} \parallel P_W^U P_F^U P_{Z_S}\right) < \tilde{\epsilon}\right) \end{aligned} \quad (185)$$

$$= 1 - \mathbb{P}_{\mathcal{B}}\left(\mathbb{D}\left(P_{WFZ_S} \parallel P_W^U P_F^U P_{Z_S}\right) < \tilde{\epsilon}, \text{ for all } S \in \mathcal{S}\right) \quad (186)$$

$$\leq 1 - \mathbb{P}_{\mathcal{B}}(\mathcal{G}) = \mathbb{P}_{\mathcal{B}}(\mathcal{G}^c) \quad (187)$$

$$\leq |\mathcal{S}||\mathcal{Z}| \exp\left(\frac{\epsilon_1^2(1 - \delta)2^\gamma}{3\tilde{W}\tilde{F}}\right), \quad (188)$$

where the inequality in (187) follows because (184) implies that

$$\mathbb{P}_{\mathcal{B}} \left(\mathbb{D} \left(P_{WFZ_S} \| P_{WP}^U P_{FS}^U \right) < \tilde{\epsilon}, \text{ for all } S \in \mathcal{S} \right) \geq \mathbb{P}_{\mathcal{B}}(\mathcal{G}). \quad (189)$$

This completes the proof for Lemma 2. The analysis in this proof is adapted from [35, Appendix].

APPENDIX C PROOF OF LEMMA 6

Let U_1, U_2, \dots, U_n be a sequence of non-negative independent random variables, which satisfy the conditions of the Lemma. For any $\theta > 0$, we have

$$\begin{aligned} \mathbb{P} \left(\sum_{i=1}^n U_i \geq (1 + \epsilon) \bar{m} \right) &= \mathbb{P} \left(e^{\theta \sum_{i=1}^n U_i} \geq e^{\theta(1+\epsilon)\bar{m}} \right) \\ &\leq \frac{\mathbb{E} \left(e^{\theta \sum_{i=1}^n U_i} \right)}{e^{\theta(1+\epsilon)\bar{m}}} \end{aligned} \quad (190)$$

$$\leq \frac{\mathbb{E} \left(e^{\theta \sum_{i=1}^n U_i} \right)}{e^{\theta(1+\epsilon)\bar{m}}} \quad (191)$$

$$= \frac{\prod_{i=1}^n \mathbb{E} \left(e^{\theta U_i} \right)}{e^{\theta(1+\epsilon)\bar{m}}} \quad (192)$$

$$\leq \frac{\prod_{i=1}^n \left(1 + \frac{e^{\theta b} - 1}{b} \mathbb{E}(U_i) \right)}{e^{\theta(1+\epsilon)\bar{m}}} \quad (193)$$

$$\leq \frac{\prod_{i=1}^n e^{\frac{e^{\theta b} - 1}{b} \bar{m}_i}}{e^{\theta(1+\epsilon)\bar{m}}} \quad (194)$$

$$\leq \frac{e^{\frac{e^{\theta b} - 1}{b} \bar{m}}}{e^{\theta(1+\epsilon)\bar{m}}} \quad (195)$$

$$= \exp \left(- \left[\theta(1 + \epsilon) - \frac{e^{\theta b} - 1}{b} \right] \bar{m} \right), \quad (196)$$

where (191) follows from Markov's inequality. (193) follows because $e^{\theta x} \leq 1 + \frac{e^{\theta b} - 1}{b} x$ for $x \in [0, b]$, as e^x is a convex function in x , (194) follows because $1 + x \leq e^x$ for all $x \geq 0$, and (195) follows because $\sum_{i=1}^n \bar{m}_i \leq \bar{m}$.

The value of θ which maximizes the right hand side of (196) is $\theta^* = \frac{1}{b} \ln(1 + \epsilon) > 0$, for which we have

$$\begin{aligned} \mathbb{P} \left(\sum_{i=1}^n U_i \geq (1 + \epsilon) \bar{m} \right) &\leq \exp \left(- \frac{\bar{m}}{b} \left[(1 + \epsilon) (\ln(1 + \epsilon) - 1) + 1 \right] \right). \end{aligned} \quad (197)$$

By considering Taylor's expansion of $x[\ln(x) - 1]$ around $x = 1$, we have, for all $x \geq 1$, that

$$x[\ln(x) - 1] + 1 \geq \frac{1}{2}(x - 1)^2 - \frac{1}{6}(x - 1)^3. \quad (198)$$

We also have, for $x \in [1, 2]$, that

$$\frac{1}{2}(x - 1)^2 - \frac{1}{6}(x - 1)^3 \geq \frac{1}{3}(x - 1)^2. \quad (199)$$

Thus, for all $x \in [1, 2]$, we have

$$x[\ln(x) - 1] + 1 \geq \frac{1}{3}(x - 1)^2. \quad (200)$$

By applying (200), with $x = (1 + \epsilon)$, to the right hand side of (197), we have, for $\epsilon \in [0, 1]$, that

$$\mathbb{P} \left(\sum_{i=1}^n U_i \geq (1 + \epsilon) \bar{m} \right) \leq \exp \left(- \frac{\bar{m}}{3b} \epsilon^2 \right). \quad (201)$$

REFERENCES

- [1] M. Nafea and A. Yener, "A new wiretap channel model and its strong secrecy capacity," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 2804–2808.
- [2] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [3] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.
- [4] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," *Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, 1984.
- [5] 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.
- [6] R. Liu, Y. Liang, H. V. Poor, and P. Spasojevic, "Secure nested codes for type II wiretap channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2007, pp. 337–342.
- [7] V. Aggarwal, L. Lai, A. R. Calderbank, and H. V. Poor, "Wiretap channel type II with an active eavesdropper," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2009, pp. 1944–1948.
- [8] M. Nafea and A. Yener, "Wiretap channel II with a noisy main channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 1159–1163.
- [9] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Semantic-security capacity for wiretap channels of type II," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3863–3879, Jul. 2016.
- [10] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.
- [11] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, Nov. 2014.
- [12] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.
- [13] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.
- [14] I. Csiszár, "Almost independence and secrecy capacity," *Problems Inf. Transmiss.*, vol. 32, no. 1, pp. 40–47, 1996.
- [15] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, May 2000, pp. 351–368.
- [16] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7377–7385, Nov. 2011.
- [17] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.
- [18] A. Gupta and S. Verdú, "Operational duality between lossy compression and channel coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3171–3179, Jun. 2011.
- [19] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
- [20] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.
- [21] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [22] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.
- [23] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
- [25] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Comput., Syst., Signal Process.*, Dec. 1984, pp. 175–179.

- [26] R. Renner, "Security of quantum key distribution," *Int. J. Quantum Inf.* vol. 6, no. 1, pp. 1–127, Feb. 2008.
- [27] M. M. Wilde, *Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
- [28] M. Tomamichel, *Quantum Information Processing With Finite Resources—Mathematical Foundations* (Springer Briefs in Mathematical Physics), vol. 5. Berlin, Germany: Springer-Verlag, 2015.
- [29] E. Tekin and A. Yener, "The Gaussian multiple access wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5747–5755, Dec. 2008.
- [30] Y. Liang and H. V. Poor, "Multiple-access channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 976–1002, Mar. 2008.
- [31] L. Lai and H. El Gamal, "The relay–eavesdropper channel: Cooperation for secrecy," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4005–4019, Sep. 2008.
- [32] X. He and A. Yener, "The role of feedback in two-way secure communications," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8115–8130, Dec. 2013.
- [33] E. Ekrem and S. Ulukus, "Capacity region of Gaussian MIMO broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5669–5680, Sep. 2012.
- [34] X. He and A. Yener, "The Gaussian many-to-one interference channel with confidential messages," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2730–2745, May 2011.
- [35] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.

Mohamed Nafea (S'11) received the B.Sc. degree in electrical engineering from Alexandria University, Alexandria, Egypt, in 2010; the M.Sc. degree in wireless communications from Wireless Intelligent Networks Center (WINC), Nile University, Giza, Egypt, in 2012, and the M.A. degree in mathematics from the Department of Mathematics, The Pennsylvania State University, University Park, PA, USA, in 2017. He is currently pursuing his Ph.D. degree in electrical engineering, and has been a Graduate Research Assistant with the Wireless Communications and Networking (WCAN) Laboratory, at The Pennsylvania State University, University Park, PA, USA, since 2012. His research interests include network information theory, information theoretic security, communication theory, statistical learning, mathematical logic and modeling theory.

Aylin Yener (S'91–M'01–SM'14–F'15) received the B.Sc. degree in electrical and electronics engineering and the B.Sc. degree in physics from Bogazici University, Istanbul, Turkey, and the M.S. and Ph.D. degrees in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ, USA. She is a Professor of Electrical Engineering at The Pennsylvania State University, University Park, PA, USA, since 2010, where she joined the faculty as an Assistant Professor in 2002. Since 2017, she is a Dean's Fellow in the College of Engineering at The Pennsylvania State University. She is currently also a Visiting Professor at the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. From 2008 to 2009, she was a Visiting Associate Professor with the same department. Her research interests include information theory, communication theory, and network science, with recent emphasis on green communications and information security. She received the NSF CAREER award in 2003, the Best Paper Award in Communication Theory in the IEEE International Conference on Communications in 2010, the Penn State Engineering Alumni Society (PSEAS) Outstanding Research Award in 2010, the IEEE Marconi Prize Paper Award in 2014, the PSEAS Premier Research Award in 2014, and the Leonard A. Doggett Award for Outstanding Writing in Electrical Engineering at Penn State in 2014.

Dr. Yener is currently a member of the Board of Governors of the IEEE Information Theory Society (2015–2020), where she was previously the treasurer (2012–2014). She served as the student committee chair for the IEEE Information Theory Society 2007–2011, and was the co-founder of the Annual School of Information Theory in North America co-organizing the school in 2008, 2009, and 2010. She was a technical (co)-chair for various symposia/tracks at the IEEE ICC, PIMRC, VTC, WCNC, and Asilomar (2005–2014). She served as an editor for the IEEE TRANSACTIONS ON COMMUNICATIONS (2009–2012), an editor and an editorial advisory board member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2001–2012), and a guest editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY (2011) and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (2015). Currently, she serves on the editorial board of the IEEE TRANSACTIONS ON MOBILE COMPUTING and as a senior editor for the IEEE JOURNAL ON SELECTED AREAS ON COMMUNICATIONS.