Finite-Blocklength Analysis in the Wiretap Channel

Bo-Yu Yang Hsuan Yu

B09901175 B09901140

Abstract

In the era of Internet of things (IoT), the tradeoff of reliability and security in the finite blocklength regime emerges as a crucial issue. The wiretap channel is one of the most typical models for characterizing this tradeoff information-theoretically. This work introduces several security metrics and delivers the core ideas about how to protect the secret message from the eavesdropper while maintaining the reliability. Moreover, both the first-order and second-order analysis for the wiretap channel are presented. We also show an inner bound of the wiretap channel with maximal leakage as the security metric, which is no larger compared to those secrecy capacity adopting mutual information, total variation secrecy, or semantic security as the security metric.

1 Introduction

With countless data being collected and transmitted over intricate networks, preserving the security through these communication schemes has become a vital issue. For example, with the advent of IoT, cloud-based services, autonomous data collecting, and online data sharing greatly increase the possible leakage of information to the third party [1, 2]. Thus, *information-theoretic security* aims at characterizing aforementioned challenges in which the security of information can be depicted by information-theoretic metrics [3, 4]. However, information-theoretic security constraints are often dependent on other design parameters, such as power consumption, latency, and reliability, which leads to unavoidable trade-offs [5]. Therefore, this motivates us to study information-theoretic security, especially the security metrics, the techniques that can enhance the security, and the trade-off between security and *reliability*, which denotes the decoding error probability can be made arbitrarily small.

To be more specific, we study the basic framework of *wiretap channel*, which is formulated in Section 1.1. Just as the channel coding in Shannon's information theory [6], he studied the fundamental limit of transmitting messages through a noisy channel reliably, which means communicating with vanishing error probability as $n \to \infty$; here, we study the first-order analysis of wiretap channel to understand how well we can guarantee on both the secrecy and reliability.

Next, we study the second-order analysis of wiretap channel. The motivation behind this kind of analysis is given below. In the era of IoT, more and more wireless devices with low-complexity terminals require low latency or short packets (e.g. traffic safety) to communicate [7, 8]. Also, it is not realistic to design privacy mechanisms with infinite blocklength, which means that we cannot let $n \to \infty$ with vanishing error probability. As a result, it is natural to investigate fundamental limits in the finite-blocklength regime.

Therefore, this report aims to convey the key concepts and strategies in existing literature to deal with the reliability-security trade-off, and to unify these seemingly varied approaches into a general recipe. This survey paper is organized as follows. We would provide the notations throughout this paper in Section 1.1. In Section 1.2, we formulate our problems, then in Section 1.3, we introduce several popular security metrics existed in literature [9, 10]. Next, in Section 2, we investigate some techniques that enhance the

security, such as *privacy amplification* [11, 12], *channel resolvability* [13–15], *random binning* [16, 17], and state their similarity. In the meantime, we would show that the reliability can be still be preserved subject to the security constraints, although when we use secrecy techniques to protect our message, the decoding difficulty for the receiver is enhanced accordingly. In Section 3, we would present the first order analysis of wiretap channel, then in Section 4, we would go through the second order analysis of wiretap channel. Section 5 provides an inner bound for the secrecy capacity with maximal leakage as the benchmark. Finally, Section 6 concludes this work by summarizing the secrecy capacity under different metrics and listing some future works.

1.1 Notations

Throughout this article, we let $\mathcal{M}$, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ denote finite sets. The cardinality of a finite set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The probability distributions on $\mathcal{X}$ is denoted as P_X. The set of probability distributions on $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$. The uniform distribution over $\mathcal{X}$ is designated as $\mathsf{Q}_{\mathcal{X}}^{\text{unif}}$. The marginal distribution of $P_X P_{Y|X}$ on Y is denoted as $P_{Y|X} \circ P_X$. The metrics between probability distribution P and Q is on a σ-algebra $\mathcal{F}$ of subsets of the set $\mathcal{X}$. The expectation is denoted as $\mathbb{E}[\cdot]$, and the variation is denoted as var. The exponential function is written as $(\cdot) \mapsto \mathsf{e}^{(\cdot)}$, and the log function is denoted as $(\cdot) \mapsto \log(\cdot)$ with base e.

The total variation distance for two probability measures P and Q on σ-algebra $\mathcal{F}$ of subsets of set $\mathcal{A}$ is defined as

$$d(\mathsf{P}, \mathsf{Q}) := \sup_{\mathcal{E} \in \mathcal{F}} |\mathsf{P}(\mathcal{E}) - \mathsf{Q}(\mathcal{E})| = \frac{1}{2} |\mathsf{P} - \mathsf{Q}|. \tag{1}$$

The Neyman-Pearson β function is denoted as

$$\beta_\alpha(\mathsf{P}, \mathsf{Q}) := \min \int \mathsf{P}_{T|X}(1|x) \mathsf{Q}(\mathrm{d}x), \tag{2}$$

where the minimization is over all random transformations $\mathsf{P}_{T|X} : \mathcal{X} \to \{0, 1\}$ satisfying

$$\int \mathsf{P}_{T|X}(1|x) \mathsf{P}(\mathrm{d}x) \geq \alpha.$$

The E_γ metric [18] is denoted as

$$\mathsf{E}_\gamma(\mathsf{P}, \mathsf{Q}) := \mathsf{P}\left[\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}} \geq \gamma\right] - \gamma \mathsf{Q}\left[\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}} \geq \gamma\right] = \sup_{\mathcal{E} \in \mathcal{F}} \left\{\mathsf{P}\left[\mathcal{E}\right] - \gamma \mathsf{Q}\left[\mathcal{E}\right]\right\}. \tag{3}$$

One can notice that when $\gamma = 1$, E_γ reduces to total variation distance. In fact, in channel resolvability problem, total variation distance is used to measure the output distribution of a channel compared to a target output distribution. However, Liu et.al [18] proposed E_γ metric to generalize total variation distance in channel resolvability problem. See Section 2.1.1 for more details.

1.2 Problem Formulation

In this project, we mainly consider the discrete memoryless wiretap channel (DM-WTC) $(W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z})$ with a sender X, a legitimate receiver Y and an eavesdropper Z, as illustrated in Fig. 1.

We assume that the message $M = (M_0, M_1)$ is uniformly distributed over the message set $\mathcal{M}_0 := [1 : 2^{n\mathrm{R}_0}]$ and $\mathcal{M}_1 := [1 : 2^{n\mathrm{R}_1}]$. A sender wants to transmit the message M to a legitimate, meanwhile ensuring that M_0 is almost independent from the eavesdropper's observation Z^n.

The average error probability constraint for the encoder and decoder is

$$\mathsf{Pr}\left[M \neq \hat{M}\right] \leq \epsilon. \tag{4}$$

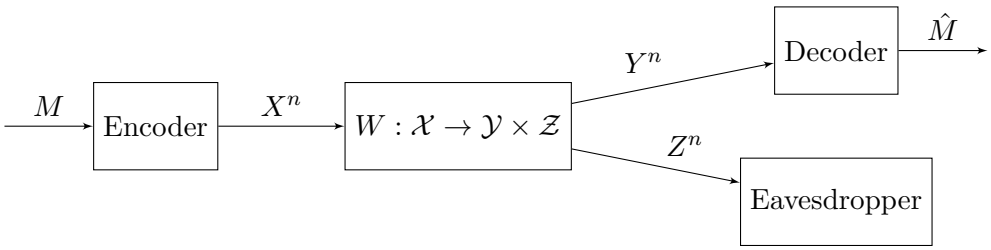

Figure 1: Wiretap channel framework

Then, the maximum error probability constraint for the encoder and decoder is

$$\mathsf{P}_{Y|M=m}\left[M \neq \hat{M}\right] \leq \epsilon. \tag{5}$$

The maximum error probability constraint is a stricter requirement than the average probability constraint, since it require every single message's decode error probability to be less than ϵ rather than overall decode error probability as in average probability constraint.

On the other hand, we hope that the secret message M_0 is almost independent from the eavesdropper's observation Z^n. Now the central question in this framework arises:

How can we characterize the information leaked to the eavesdropper through this wiretap channel while maintaining the reliability?

There are several candidates as security metrics to characterize the information leakage rate, such as mutual information, total variation secrecy, distinguishing security, semantic security, and maximal leakage. The definitions and relations of these benchmarks are summarized in the upcoming subsection.

1.3 Security Metrics

1.3.1 Mutual Information Secrecy

Since mutual information depicts dependency between two distributions, it is natural to define secrecy by Shannon's mutual information. The corresponding definition of secrecy is listed below.

Definition 1 (Weak Mutual Information Secrecy [4])**.** *Suppose that M is uniformly distributed, then the weak secrecy of message M and eavesdropper's observation Z^n is defined as*

$$\mathrm{I}^{\mathsf{weak}}(M; Z^n) := \frac{1}{n}\mathrm{I}(M; Z^n)\,. \tag{6}$$

However, assuming messages are uniformly distributed is unrealistic, so, in 2000, Maurer and Wolf [19] further provide the definition of strong secrecy.

Definition 2 (Strong Mutual Information Secrecy [4])**.** *The strong secrecy of message M and eavesdropper's observation Z^n is defined as*

$$\mathrm{I}^{\mathsf{strong}}(M; Z^n) := \mathrm{I}(M; Z^n)\,. \tag{7}$$

Note that strong secrecy is just mutual information between message M and eavesdropper's observation Z^n.

1.3.2 Total Variation Secrecy

The ideal scenario of wiretap channel is that the message M is not only uniformly distributed but also independent of eavesdropper's observation Z. In this scenario, eavesdropper literally learns nothing from the channel, and hence the privacy requirement is satisfied. Therefore, total variance secrecy captures the *distance* of the wiretap channel to the ideal scenario mentioned above.

Definition 3 (Average Total Variation Secrecy [20, 21])**.**

$$S(M|Z) := d(\mathsf{P}_{MZ}, \mathsf{Q}_{\mathcal{M}}^{unif}\mathsf{P}_Z). \tag{8}$$

Average total variation secrecy measures the *overall* performance of message M and eavesdropper's observation Z. In the meantime, one can consider the secrecy of a specific message $M = m$ and the $\mathsf{P}_{Z|M=m}$ induced by that specific message. Inspired by this observation, Yang et al. [10] proposed maximal total variation secrecy, whose definition is given below.

Definition 4 (Maximal Total Variation Secrecy [10])**.**

$$S^{\max}(M|Z) := \max_{m\in\mathcal{M}} d(\mathsf{P}_{Z|M=m}, \mathsf{Q}_Z), \tag{9}$$

where $\mathsf{Q}_Z = P_{Z|M} \circ \mathsf{Q}_{\mathcal{M}}^{unif}$.

Note that maximal total variation secrecy is stricter than average total variation secrecy since it requires every message $M = m$ inducing $\mathsf{P}_{Z|M=m}$ being similar to Q_Z; on the contrary, average total variation secrecy only requires the overall security performance.

1.3.3 Distinguishing Security and Semantic Security

Now we provide secrecy metrics widely used in cryptography literature, which is *distinguishing security* and *semantic security*. Then, we state two theorems that compare the mutual information, total variation, distinguishing security and semantic security.

Definition 5 (Distinguishing Security [9, 22])**.** *For the message $M = m_0$ and $M = m_1$, the distinguishing security is defined as*

$$\mathsf{DS}(M|Z) := \max_{m_0,m_1\in\mathcal{X}} d(\mathsf{P}_{Z|M=m_0}, \mathsf{P}_{Z|M=m_1}). \tag{10}$$

Definition 6 (Semantic Security [9, 22])**.** *Let $M = f(X)$, and thus $M - X - Z - \hat{M}$ forms a Markov chain. The eavesdropper's inference of M after observing Z is $\hat{m}(Z)$. The semantic security is defined as*

$$\mathsf{SS}(M|Z) = \sup_{M:M-X-Z} (\sup_{\hat{m}(\cdot)} \mathsf{Pr}\{M = \hat{m}(Z)\} - \max_{m\in M} \mathsf{P}_M(m)). \tag{11}$$

Note that we can related the above mentioned security metrics by the theorems below.

Theorem 1 (Mutual Information Secrecy and Distinguishing Security [9])**.** *For every transition probability distribution $\mathsf{P}_{Z|M} : \mathcal{M} \to \mathcal{Z}$*

$$\frac{\mathsf{DS}(M|Z)^2}{2} \le \mathrm{I}^{\mathsf{strong}}(M; Z^n) \le 2\mathsf{DS}(M|Z) \log \frac{2^n}{\mathsf{DS}(M|Z)}. \tag{12}$$

One can refer to [9, Thm. 5 and 8] for the full proof.

Theorem 2 (Total Variation Secrecy, Distinguishing Security, and Semantic Security [10]). *For every transition probability distribution* $\mathsf{P}_{Z|M} : \mathcal{M} \to \mathcal{Z}$

$$\mathsf{SS}(M|Z) \leq S^{\max}(M|Z) \leq \mathsf{DS}(M|Z) \leq 2\mathsf{SS}(M|Z) \leq 2S^{\max}(M|Z). \tag{13}$$

One can refer to [9, Theorem 1] and [10, Theorem 1] for the proof.

Note that the above two theorems help us to relate the secrecy metrics used in information theory to security metrics used in cryptography literature.

1.3.4 Maximal Leakage

In 2018, Issa et al. proposed a measure, called maximal leakage, to operationally characterize how much information an observation Z leaks through a side channel about the sensitive information X. Moreover, maximal leakage satisfies some axiomatic properties, such as data-processing inequality, independence property, and additivity. Here we recapitulate the definition of the maximal leakage.

Definition 7 (Maximal Leakage [23]). *Let $M = f(X)$ such that $M - X - Z - \hat{M}$ form a Markov chain. The eavesdropper's inference of M after observing Z is $\hat{m}(Z)$. The maximal leakage is defined as*

$$\mathcal{L}_{\infty}^{\max}(X \to Z) := \sup_{M-X-Z-\hat{M}} \log \frac{\mathsf{Pr}\{M = \hat{m}(Z)\}}{\max_{m \in M} \mathsf{P}_M(m)}. \tag{14}$$

Issa et al. also showed that the maximal leakage from X to Z is given by the Sibson mutual information of order infinity.

Proposition 3 (Characterization of Maximal Leakage [23]). *For any joint distribution P_{XZ} on finite alphabets $\mathcal{X}$ and $\mathcal{Z}$, the maximal leakage from X to Z can be depicted as*

$$\mathcal{L}_{\infty}^{\max}(X \to Z) = \log \sum_{z \in \mathcal{Z}} \max_{\substack{x \in \mathcal{X}: \\ \mathsf{P}_X(x) > 0}} \mathsf{P}_{Y|X}(y|x) = \mathrm{I}_{\infty}^{\mathrm{S}}(X;Z)\,, \tag{15}$$

where Sibson mutual information of order infinity and max-divergence are defined as

$$\mathrm{I}_{\infty}^{\mathrm{S}}(X;Z) := \inf_{\mathsf{Q}_Z} \mathrm{D}_{\infty}(\mathsf{P}_{XZ} \| \mathsf{P}_X \times \mathsf{Q}_Z)\,; \tag{16}$$

$$\mathrm{D}_{\infty}(\mathsf{P} \| \mathsf{Q}) := \max_{x \in \mathsf{supp}(\mathsf{Q}_X)} \log \frac{\mathsf{P}_X(x)}{\mathsf{Q}_X(x)}. \tag{17}$$

The semantic security introduced previously considers the difference between the prior and posterior guessing probabilities. Sometimes, the messages M of interest such as passwords are hard to guess (i.e. $\max_m \mathsf{P}_M(m)$ is small), the ratio between the guessing probabilities is arguably the more appropriate measure of the change. Here, we give an example to illustrate this motivation to adopting the maximal leakage as the security metric.

Example 1. *Consider $M = X$ and let X be uniformly distributed on $\mathcal{X}$. We define the random variable Z as*

$$Z = \begin{cases} X, \text{ if } X \bmod 8 = 0, \\ 1, \text{ otherwise.} \end{cases}$$

Then, the semantic security and maximal leakage from X to Z in this example can be respectively derived as below:

$$\mathsf{SS}(M|Z) = \frac{1}{8} + \frac{1}{2^{8n}} - \frac{1}{2^{8n}} = \frac{1}{8};$$
$$\mathcal{L}_{\infty}^{\max}(X \to Z) = \log(2^{8n-3} + 1).$$

We can observe that maximal leakage serves as a stricter security constraint, and it depends on blocklength n in this scenario.

2 Tradeoff between Reliability and Security

In this chapter, we first summarize and compare several techniques to analyze the security constraint in wiretap channels, and then illustrate a unified high-level approach to investigate the reliability-security trade-off.

2.1 A Toolbox for Analyzing Security

Here we introduce the two popular building blocks that can enhance security: channel resolvability and privacy amplification.

2.1.1 Channel Resolvability

Channel resolvability [10, 13–15, 18] refers to the minimum randomness rate needed for an input process to approximate the channel output distribution. To elaborate, given a stationary memoryless channel, $W_1 : \mathcal{X} \to \mathcal{Z}$ and an input distribution P_X inducing output distribution $\mathsf{P}_Z = \mathsf{P}_X \circ \mathsf{P}_{W_1}$, it is possible to well approximate the output product distribution P_Z^n with n independent uses of channel W_1.

When it comes to wiretap channels, we can leverage channel resolvability to characterize how much information is needed for equivocating the eavesdropper. In other words, we would like to design a coding scheme to simulate a side channel with target distribution Q_Z^n being almost independent to the secret message M_0 to confuse the eavesdropper and hence the leaked information between (M, Z^n) through the wiretap channel is restrained. Given a resolvability code $\mathcal{C}_n = \{x_1^n, ..., x_{|\mathcal{M}|}^n\}$ of blocklength n and size $|\mathcal{M}|$, when a uniformly chosen codeword from $\mathcal{C}_n$ is transmitted, we denote the output distribution of $W_1 : \mathcal{X} \to \mathcal{Z}$ as

$$\mathsf{P}_{\mathcal{C}_n}(z^n) := \frac{1}{|\mathcal{M}|} \sum_{i=1}^{|\mathcal{M}|} W_1^n(z^n|x_i^n).$$

In the realm of wiretap channel, we concern the independency between the secret message M_0 and eavesdropper's observation Z^n. To measure this behavior quantitatively, we can set the target distribution in the form of $\mathsf{P}_{M_0}\mathsf{Q}_{Z^n}$ and measure $\mathbb{D}(\mathsf{P}_{M_0 Z^n}, \mathsf{P}_{M_0}\mathsf{Q}_{Z^n})$, where $\mathbb{D}(\cdot,\cdot)$ denotes a general distance measure. There are several distance criteria $\mathbb{D}(\cdot,\cdot)$ to measure how close the simulated output distribution is to the target distribution, such as variational distance (e.g. E_γ metrics) [18, 24], KL divergence [25], and Rényi divergence [14, 15]. In order to approximate the message-observation-independent distribution $\mathsf{P}_{M_0}\mathsf{Q}_{Z^n}$, the distance between our simulated channel output $\mathsf{P}_{M_0 Z^n}$ and the target $\mathsf{P}_{M_0}\mathsf{Q}_{Z^n}$ should be arbitrarily small as the blocklength n increases. Later in Section 5, we will show that an inner bound of the secrecy capacity for the wiretap channel with maximal leakage can be derived with the aid of channel resolvability with Rényi divergence as the distance metric:

Proposition 4 (Rényi resolvability [14])**.** *Let $M \mapsto f_{\mathcal{C}}(M)$ be a random function induced by $\mathcal{C}$. For $s = (0, 1] \cup \{\infty\}$, we have*

$$\inf\left\{\mathrm{R} : \inf_{f_{\mathcal{C}}} \mathrm{D}_{1+s}(\mathsf{P}_{\mathcal{C}Z^n} \| \mathsf{P}_{\mathcal{C}}\mathsf{Q}_{Z^n}) \to 0\right\} = \min_{\mathsf{P}_X \in \mathcal{P}(\mathsf{P}_{Z|X}, \mathsf{Q}_Z)} \sum_x \mathsf{P}_X(x) \mathrm{D}_{1+s}\left(\mathsf{P}_{Y|X}(\cdot|x) \middle\| \mathsf{Q}_{Z^n}\right) \tag{18}$$

$$=: \mathrm{R}_{1+s}(\mathsf{P}_{Z|X}, \mathsf{Q}_Z). \tag{19}$$

2.1.2 Privacy Amplification

This subsection is mainly contributed by celebrated works in [11, 26, 27], one can also refers to [12, 28] if needed.

Privacy amplification aims to distill highly secret information from a *partially* secret source, where the eavesdropper have access to it. The goal is to convert these kind of sources to a new code that is uniformly distributed and nearly independent of the eavesdropper's observation. To describe it mathematically, we should design a good mapping function $g(X) : \mathcal{X} \to \mathcal{M}$ for the source X such that $M = g(X)$ is uniformly distributed and independent of eavesdropper's observation Z. In cryptography literature, $g(\cdot)$ is usually referred to *hashing* function, but in information theory, $g(\cdot)$ is called a *code*.

To achieve privacy amplification, Bennett et al. [11] uses the technique *universal hashing* introduced in [29]. Particularly, for a set of functions $G : \mathcal{X} \to \mathcal{M}$ being *universal*, it must satisfy that for every function $g \in G$,

$$\mathsf{Pr}\left[g(x_1) = g(x_2)\right] \leq \frac{1}{|\mathcal{M}|}, \forall\ x_1 \neq x_2. \tag{20}$$

There is a beautiful lemma, called *leftover hash lemma*, which plays a significant role in proving the performance of universal hashing. We state the lemma and give a sketch of proof below.

Lemma 1 (Leftover Hash Lemma [12])**.** *For universal hash function sets G, then for every P_{XZ} and every Q_Z supported on $\mathcal{Z}$, we have*

$$\mathbb{E}_G\left[S(G|\mathsf{P}_{X|Z})\right] \leq \frac{1}{2}\sqrt{|\mathcal{M}|\mathsf{e}^{-H_2(\mathsf{P}_{XZ}|\mathsf{Q}_Z)}}. \tag{21}$$

Proof. (Sketch of proof) By expanding the term of the left-hand side, we have

$$\mathbb{E}_G\left[S(G|\mathsf{P}_{XZ})\right] = \mathbb{E}_G\left[d(\mathsf{P}_{G(X)Z}|\mathsf{Q}_{\mathcal{M}}^{\text{unif}}\mathsf{P}_Z)\right] \tag{22}$$

$$= \frac{1}{2}\sum_{m\in\mathcal{M},z\in\mathcal{Z}} \mathbb{E}_G\left[|\sum_{x\in G^{-1}(x)} \mathsf{P}_{XZ}(x,z) - \frac{1}{|\mathcal{M}|}\mathsf{P}_Z(z)|\right] \tag{23}$$

$$= \frac{|\mathcal{M}|}{2}\sum_{z\in\mathcal{Z}} \mathbb{E}_G\left[|\sum_{x\in G^{-1}(1)} \mathsf{P}_{XZ}(x,z) - \frac{1}{|\mathcal{M}|}\mathsf{P}_Z(z)|\right]. \tag{24}$$

The last equation (24) comes from the symmetry of set G. Note that since

$$\mathbb{E}\left[\sum_{x\in G^{-1}(1)} \mathsf{P}_{XZ}(x,z)\right] = \frac{\mathsf{P}_Z(z)}{|\mathcal{M}|}, \tag{25}$$

we can define $A(z) = \sum_{x\in G^{-1}(1)} \mathsf{P}_{XZ}(x,z)$, and rewrite the last equation (24) as

$$\frac{|\mathcal{M}|}{2}\sum_{z\in\mathcal{Z}} \mathbb{E}_G\left[|A(z) - \mathbb{E}\left[A(z)\right]|\right] = \frac{|\mathcal{M}|}{2}\sum_{z\in\mathcal{Z}} \mathbb{E}_G\left[\sqrt{\mathsf{Var}(A(z))}\right]. \tag{26}$$

Since there are second-order terms of $P_{XZ}(x,z)$ in $\mathsf{Var}(A(z))$ as we defined before, we can related to second-order of Rényi entropy and conditional Rényi entropy defined as below.

The second-order Rényi entropy is defined as

$$\mathrm{H}_2\left(\mathsf{P}_{XZ}\right) = -\log \sum_{x\in\mathcal{X},z\in\mathcal{Z}} \mathsf{P}_{XZ}(x,z)^2; \tag{27}$$

whereas the conditional second-order of Rényi entropy is defined as

$$\mathrm{H}_2\left(\mathsf{P}_{XZ}|\mathsf{Q}_Z\right) = -\log \sum_{x\in\mathcal{X},z\in\mathcal{Z}} \frac{\mathsf{P}_{XZ}(x,z)^2}{\mathsf{Q}_Z(z)}. \tag{28}$$

Then, the main intuition in the proof of leftover hash lemma is that observing the term $H_2(G(X))$ is

$$\mathrm{H}_2(G(X)) = -\log \sum_{x \in G^{-1}(x)} \mathsf{P}_X(x)^2, \tag{29}$$

and that $\mathsf{P}_X(x)^2 = \mathsf{Pr}\left[G(X_1) = G(X_2)\right]$ is the *collision* probability when $G(X_1)$ and $G(X_2)$ take on the same value. Then, we can observe that

$$\mathbb{E}_G\left[\mathsf{e}^{-\mathrm{H}_2(G(X))}\right] = \mathsf{Pr}\left[G(X_1) = G(X_2)\right] \tag{30}$$

$$\leq \mathsf{Pr}\left[X_1 = X_2\right] + \mathsf{Pr}\left[G(X_1) = G(X_2) | X_1 \leq X_2\right] \tag{31}$$

$$= \mathsf{e}^{\mathrm{H}_2(X)} + \frac{1}{|\mathcal{M}|}. \tag{32}$$

Note that the similar bound holds for $\mathrm{H}_2(\mathsf{P}_{XZ}|\mathsf{Q}_Z)$. Then, just as in proof in [12, Sec 5.5], using Cauchy-Schwarz inequality to lift the l_1 distance of $\mathsf{P}_{G(X)Z}$ and $\mathsf{Q}_{\mathcal{M}}^{\mathrm{unif}}\mathsf{P}_Z$ to l_2 distance, we can conclude the proof. □

2.2 A General Approach to Reliability-Security Trade-offs

Recall that for a wiretap channel, we hope that the reliability and security can be guaranteed at the same time.

On the one hand, the reliability can usually be proved by the standard random coding argument or dependent testing bound. We would state two common used bounds for reliability, which is random coding union bound and dependent testing bound. Then, we would sketch how the decoding parts are done. First of all, both encoders for random union bound and dependent testing bound use Shannon's random coding scheme [6] as taught in class. However, for the decoding part, both bounds use different method. For random coding union bound, the reliability for the average decoding error probability and maximal decoding error probability are given below.

Lemma 2 (Random coding union bound [30])**.** *The random coding union bound for average error probability for message M is*

$$\epsilon_{\mathsf{RCU}}(M) = \mathbb{E}\left[\min\{1, (|\mathcal{M}|-1)\mathsf{Pr}\left[i(\bar{X};Y) \geq i(X;Y)|X,Y\right]\}\right], \tag{33}$$

where $i(x;y) = \log \frac{\mathsf{dP}_{Y|X}(y|x)}{\mathsf{dP}_Y(y)}$ and $\mathsf{P}_{XY\bar{X}}(x,y,\bar{x}) = \mathsf{P}_X(x)\mathsf{P}_{Y|X}(y|x)\mathsf{P}_{\bar{X}}(\bar{x})$. The random coding union bound for maximal error probability for message M is

$$\epsilon_{\mathsf{RCU},\max}(M) = \inf_{\tau \in (0,1)} \{\tau^{-1} \frac{\epsilon_{\mathsf{RCU}}((1-\tau)M)}{\tau}\}. \tag{34}$$

For the proof of random coding union bound, one can refers to [30, Thm. 16, eq. (220)]. The main idea is that, the decoder uses the *maximal likelihood* decoding technique to decode. Thus, the error probability $\mathsf{Pr}\left[\cdot\right]$ terms refer to that for the probability that under condition X, Y, there exists an $\bar{X}, Y$ whose probability is greater than XY. In this scene, the maximal likelihood decoder would decode a wrong message, since except the correct one, there are $|\mathcal{M}|-1$ messages, the probability is multiplied by the factor of $|\mathcal{M}|-1$.

For dependent testing bound, the reliability analysis for the average decoding error probability and maximal decoding error probability are given below.

Lemma 3 (Dependent testing bound [30])**.** *The dependent testing bound for average error probability for message M is*

$$\epsilon_{\mathsf{DT}}(\frac{|\mathcal{M}|-1}{2}) = 1 - E_{\frac{|\mathcal{M}|-1}{2}}(\mathsf{P}_{XY}, \mathsf{P}_X\mathsf{P}_Y), \tag{35}$$

where for $\gamma = \frac{|\mathcal{M}|-1}{2}$*, the definition of* E_γ *metric is provided in equation (3). The dependent testing bound for maximal error probability for message* M *is*

$$\epsilon_{\mathsf{DT},\max}(M) = \inf_{\gamma>0}\{\mathsf{Pr}\left[i(X;Y) \leq \log\gamma\right] + |\mathcal{M}| \sup_{x\in\mathcal{X}} \mathsf{Pr}\left[i(x;Y) \geq \log\gamma\right]\}, \tag{36}$$

with $i(x;y) = \log \frac{\mathsf{dP}_{Y|X}(y|x)}{\mathsf{dP}_Y(y)}$.

For the proof of dependent testing bound (also the reason why $\gamma = \frac{|\mathcal{M}|-1}{2}$), one can refers to [30, Thm. 17, Thm. 21]. The main idea is that, the decoder uses *threshold* decoder which is equivalent to Feinstein's suboptimal decoder [31]. To be more specific, the decoder can treat $|\mathcal{M}|$ messages as $|\mathcal{M}|$ likelihood ratio binary hypothesis tests in parallel with threshold γ. Then, it decodes according to the likelihood ratio of the codewords. One can observe the similar mathematical structure in the likelihood ratio hypothesis testing.

On the other hand, for the security part, what we concern is how to design a code to restrain the eavesdropper from guessing the secret message M_0 correctly. The most ideal case is that all messages (M_0, M_1) delivered through the wiretap channel are totally independent of the eavesdropper's observation Z^n . However, there is no free lunch. To obtain the security guarantee, we must sacrifice some transmission rate to protect our secret message M_0. But how much ratio of communication rate should we distribute to protect our secret data? Put more precisely, how much information is required to let the secret message M_0 be almost independent to eavesdropper's observation Z^n?

To answer this question, we can utilize any of the two security building blocks introduced in the previous subsection. Channel resolvability investigates how much information are needed to simulate the target output distribution. The minimum rate for achieving this criterion is exactly what we want to characterize. For another, privacy amplification allows us to convert an arbitrary non-secret code M into a secrecy code M_0. If we can extract and amplify the intrinsic randomness in our message, in the eavesdropper's viewpoint, it becomes difficult to decode the seemingly uniform message. In fact, it can be shown that channel resolvability and privacy amplification are based on the same probabilistic principle [10]. Lemma 4 and 5 showcase that these two strategies admit upper bounds of the same form with total variation distance being the security metric.

Lemma 4 (Upper bound by privacy amplification [10])**.** *For any* $\gamma > 0$ *and probability distribution* $\mathsf{Q}_Z \in \mathcal{P}(\mathcal{Z})$

$$S(G|Z) \leq \mathsf{E}_\gamma(\mathsf{P}_{XZ}, \mathsf{Q}_{\mathcal{M}}^{\text{unif}}\mathsf{Q}_Z) + \frac{1}{2}\sqrt{\frac{\gamma}{L}\mathbb{E}_{\mathsf{P}_{XZ}}\left[\exp\left(-\left|\iota(X;Z) - \log\gamma\right|\right)\right]}, \tag{37}$$

where

$$L := \frac{|\mathcal{X}|}{|\mathcal{M}|}; \tag{38}$$

$$\iota(X;Z) := \log \frac{\mathrm{d}\mathsf{P}_{XZ}}{\mathrm{d}(\mathsf{Q}_{\mathcal{M}}^{\text{unif}}\mathsf{Q}_Z)}(x,z). \tag{39}$$

One can refer to [10, Lemma 2] for the proof. One might wonder the difference of the bound on leftover hash lemma in equation (21) and the one present here. The key difference is that the one presented here partitions the distribution P_{XZ} into non-typical parts A and typical parts B. Then, it only performs leftover hash lemma in the typical parts B. The reason for doing this is simple, since for non-typical parts A, using leftover hash lemma to increase secrecy require extra bits. However, non-typical parts are unlikely to happen asymptotically. Therefore, we only need to apply leftover hash lemma on the typical parts B. By triangle inequality, we have

$$\mathbb{E}\left[\left|A + B - \mathbb{E}\left[A+B\right]\right|\right] \leq 2\mathbb{E}\left[A\right] + \mathbb{E}\left[\left|B - \mathbb{E}\left[B\right]\right|\right]. \tag{40}$$

Then, after applying leftover hash lemma in second term [12], and for every probability distribution R_{XZ}, we have

$$\mathbb{E}_{\mathbb{G}}\left[S(G|\mathsf{P}_{XZ})\right] \leq |\mathsf{P}_{XZ} - \mathsf{R}_{XZ}| + \frac{1}{2}\sqrt{M\mathsf{e}^{-H_2(\mathsf{R}_{XZ}|\mathsf{Q}_Z)}}. \tag{41}$$

As in the proof in [10, Lemma. 2], the minimizer of the above equation is

$$\mathsf{R}^*_{XZ}(x,z) = \begin{cases} \mathsf{P}_{XZ}(x,z), & \iota(X;Z) \leq \log\gamma; \\ \frac{\gamma}{|\mathcal{X}|}\mathsf{Q}_Z(z), & \text{otherwise.} \end{cases}$$

By plugging $\mathsf{R}^*_{XZ}(x,z)$ into equation (41), we can get the upper bound presented in equation (37).

Lemma 5 (Upper bound by channel resolvability [10])**.** *Let $\mathcal{C} := X_1, ..., X_L$ denote a random codebook whose codewords are independently and identically generated from P_X. For any $\gamma > 0$ and $\mathsf{Q}_Z \in \mathcal{P}(\mathcal{Z})$,*

$$\mathbb{E}\left[d(\mathsf{P}_{Z|\mathcal{C}}, \mathsf{P}_Z)\right] \leq \mathsf{E}_\gamma(\mathsf{P}_{XZ}, \mathsf{Q}^{\text{unif}}_{\mathcal{M}}\mathsf{Q}_Z) + \frac{1}{2}\sqrt{\frac{\gamma}{L}\mathbb{E}_{\mathsf{P}_{\tilde{X}Z}}\left[\exp\left(-\left|\iota(\tilde{X};Z) - \log\gamma\right|\right)\right]},$$

where the expectation is taken with respect to $(\tilde{X}, Z) \sim \mathsf{Q}^{\text{unif}}_{\mathcal{M}}\mathsf{P}_{Z|X}$, L and $\iota(X;Z)$ are defined in (38) and (39).

The proof is in [10, Lemma. 4].

Remark 1. One may wonder why channel resolvability and privacy amplification obtain the same upper bound. Our interpretation is that first of all, both techniques aim at choosing good codewords to achieve the ideal scenario, or the perfect secrecy that the eavesdropper's observation Z is independent from X and the message M is uniformly distributed. (Note that $g(X)$ is referred to *code* in information theory, as we mentioned in subsection 2.1.2.) Also, it is not hard to see that if we set $|\mathcal{M}| = 2^{nR}$, the privacy amplification is equivalent to *random binning* taught in class. Besides, both techniques highly depends on P_X, since for channel resolvability, it aims to find a good codebook generated from P_X to achieve ideal output distribution, and privacy amplification concerns constructing a good mapping, or random binning of source P_X to the similar output distribution to confuse the eavesdropper. Therefore, we can expect that the bound would in the similar form. In conclusion, channel resolvability and privacy amplification highly depend on probability distribution P_X and use the same probability argument, so their upper-bounds can achieve the same form.

Apart from allocating the sufficient amount of transmission rate to protect the secret message, we can further trade reliability for secrecy [10, 25]. If we continue adopting total variation distance (equivalent to mutual information and semantic security [9]) as the security metric, the following theorem indicates that reliability can be traded for secrecy proportionally.

Theorem 5 (Trading reliability for secrecy [10])**.** *Let ϵ, ϵ_0 denote the error probability tolerance and δ, δ_0 denote the total variation secrecy tolerance, and n be the blocklength. The maximal secrecy rate is defined as*

$$\mathrm{R}^*(n,\epsilon,\delta) := \max\left\{\frac{\log|\mathcal{M}|}{n}\middle|\exists \text{ a secrecy code s.t. both the error prob. and secrecy constraint are met}\right\}.$$

For $\epsilon > \epsilon_0$ and $\delta = \delta_0\frac{1-\epsilon}{1-\epsilon_0}$, we have

$$\mathrm{R}^*(n,\epsilon_0,\delta_0) \leq \mathrm{R}^*(n,\epsilon,\delta) = \mathrm{R}^*\left(n,\epsilon,\delta_0\frac{1-\epsilon}{1-\epsilon_0}\right). \tag{42}$$

3 First-Order Analysis

In general, the first order secrecy capacity of a discrete memoryless wiretap channel with weak mutual information secrecy can be characterized by the following theorem.

Theorem 6 (The first-order secrecy capacity with weak MI secrecy [16, 32])**.** *For $|\mathcal{U}| \leq |\mathcal{X}|$, the secrecy capacity of the DM-WTC is*

$$\mathrm{C_S} = \max_{\mathsf{P}_{UX}} \mathrm{I}(U;Y) - \mathrm{I}(U;Z)\,. \tag{43}$$

If the channel $W_1 : \mathcal{X} \to \mathcal{Z}$ becomes the degraded version $W_1^{\mathrm{d}} : \mathcal{Y} \to \mathcal{Z}$ with $\mathsf{P}_{Y,Z|X} = \mathsf{P}_{Y|X}\mathsf{P}_{Z|Y}$, then the secrecy capacity simplifies to

$$\mathrm{C_S} = \max_{\mathsf{P}_X} \mathrm{I}(X;Y) - \mathrm{I}(X;Z)\,. \tag{44}$$

Remark 2. To obtain a secrecy communication rate strictly greater than zero, the channel W_0 should be less noisy than the channel W_1. This is due to the "less-noisy" condition can be expressed as $\mathrm{I}(X;Y|U) - \mathrm{I}(X;Z|U) \geq 0$. We have

$$\mathrm{I}(U;Y) - \mathrm{I}(U;Z) = \mathrm{I}(X;Y) - \mathrm{I}(X;Z) - (\mathrm{I}(X;Y|U) - \mathrm{I}(X;Z|U)) \overset{\text{(a)}}{\leq} \mathrm{I}(X;Y) - \mathrm{I}(X;Z)\,, \tag{45}$$

where (a) holds if and only if $\mathrm{I}(X;Y|U) - \mathrm{I}(X;Z|U) \geq 0$, i.e. channel W_1 is more capable than W_0.

An upper bound of (43) is given by [33, Thm. 1]

$$\mathrm{C_S} \leq \max_{\mathsf{P}_X} \mathrm{I}(X;Y|Z)\,, \tag{46}$$

which holds for a WTC with public feedback. The upper bound is tight if there exists a unique probability distribution that attains maximum in (46). Also, the strong converse in (46) holds when the wiretap channel is degraded [10, 33].

Proof sketch of Theorem 6

Recall that a wiretap channel should satisfy both reliability and security.

On the one hand, suppose that both reliability and security have been satisfied. The converse part can be proved by applying Fano's inequality first, and then using Csiszár's sum identity together with auxiliary random variables to help single-letterize the expression.

On the other hand, the direct part can be proved with the aid of random binning. Next, we will give a proof sketch to show that the secrecy capacity in (43) can satisfy both requirements if the rate pair $(\mathrm{R}_0, \mathrm{R}_1)$ meets

$$\begin{cases} \mathrm{R}_0 \leq \max\limits_{\mathsf{P}_{UX}} \mathrm{I}(U;Y) - \mathrm{I}(U;Z)\,; \\ \mathrm{R}_1 \geq \mathrm{I}(U;Z)\,. \end{cases}$$

Reliability.

This part can be proved using standard random coding argument. If $\mathrm{R}_0 + \mathrm{R}_1 \leq \mathrm{I}(U;Y) - \delta(\epsilon)$, $\overline{\mathsf{P}}_{\mathsf{e}} \to 0$ as $n \to \infty$ by the law of large numbers (LLN) and packing lemma.

Security.

Let l denote the index randomly selected by the encoder. Then every codebook $\mathcal{X}$ induces a conditional pmf of the form $\mathsf{P}(m,l,u^n,z^n) = 2^{-n\mathrm{R}_0}2^{-n\mathrm{R}_1}\mathsf{P}(u^n|l,\mathcal{C})\prod_{i=1}^{n}\mathsf{P}_{Z|U}(z|u)$. To prove that the information leakage rate with weak mutual information secrecy is bounded, we can upper-bound the equivocation rate $\mathrm{H}(M|Z^n)$ by Chebyshev inequality and LLN over each random assignment $\mathcal{C}$. Specifically, if $\mathrm{R}_0+\mathrm{R}_1 < \mathrm{I}(U;Y)-\delta(\epsilon)$, it can be shown that

$$\begin{aligned}\mathrm{I}(M;Z^n|\mathcal{C}) &= \mathrm{H}(M|C) - \mathrm{H}(M|Z^n,\mathcal{C}) && (47)\\ &\overset{\text{(a)}}{=} n\mathrm{R}_0 - (\mathrm{H}(L|Z^n,\mathcal{C}) - \mathrm{H}(L|Z^n,M,\mathcal{C})) && (48)\\ &\overset{n\to\infty}{\leq} n\delta(\epsilon), && (49)\end{aligned}$$

where (a) follows from $M \overset{\text{(f)}}{=} L$. Therefore, the weak secrecy is guaranteed.

For the complete version of the proof, please refer to [16, Chap. 22].

4 Second-Order Analysis

Remind that we have defined average error probability (Eq. (4)) and maximum error probability (Eq. (5)) until now; also, we have defined average total variation secrecy (Eq. (8)) and maximal total variation secrecy (Eq. (9)). Therefore, we can formulate average secrecy rate problem and maximal secrecy problem with these constraint as following respectively.

For a code $(n,|\mathcal{M}|,\epsilon,\delta)^{\text{avg}}$ satisfying the average error probability constraint $\mathsf{Pr}\left[M\neq\hat{M}\right]\leq\epsilon$ and the average total variation secrecy constraint $S(M|Z)\leq\delta$, $\mathrm{R}^{\text{avg}}(n,\epsilon,\delta)$ is defined as

$$\mathrm{R}^{\text{avg}}(n,\epsilon,\delta) := \max\left\{\frac{\log|\mathcal{M}|}{n} : \exists(n,|\mathcal{M}|,\epsilon,\delta)^{\text{avg}} \text{ secrecy code}\right\}. \tag{50}$$

Analogously, for a code $(n,|\mathcal{M}|,\epsilon,\delta)^{\max}$ satisfies the maximal error probability constraint as $\mathsf{P}_{Y|M=m}\left[M\neq\hat{M}\right]\leq\epsilon$ and the maximal total variation secrecy constraint $S^{\max}(M|Z)\leq\delta$, the rate $\mathrm{R}^{\max}(n,\epsilon,\delta)$ is defined as

$$\mathrm{R}^{\max}(n,\epsilon,\delta) := \max\left\{\frac{\log|\mathcal{M}|}{n} : \exists(n,|\mathcal{M}|,\epsilon,\delta)^{\max} \text{ secrecy code}\right\}. \tag{51}$$

It is not hard to see that $\mathrm{R}^{\text{avg}}(n,\epsilon,\delta)\leq\mathrm{R}^{\max}(n,\epsilon,\delta)$, since the maximal secrecy rate's requirement on reliability and secrecy is stricter than the average secrecy rate's. Thus, one might wonder what is the lower bound that is achievable for maximal secrecy rate $\mathrm{R}^{\text{avg}}(n,\epsilon,\delta)$ and the converse upper bound on average secrecy rate $\mathrm{R}^{\max}(n,\epsilon,\delta)$. Here we provide the results from the work in [10, Thm. 13], which is the latest results to the best of our knowledge. For both achievability and converse, we would provide sketch of proofs and indicate the crucial steps.

4.1 Achievability

The lower bound of maximal secrecy rate is given as below.

Theorem 7. *For a code $(n,|\mathcal{M}|,\epsilon,\delta)^{\max}$ with $\epsilon+\delta<1$, we have*

$$\mathrm{R}^{\max}(n,\epsilon,\delta) \geq \mathrm{C}_S - \frac{V_1}{n}Q^{-1}(\epsilon) - \frac{V_2}{n}Q^{-1}(\delta) + \mathcal{O}\left(\frac{\log n}{n}\right), \tag{52}$$

where for P_X *and* $\mathsf{P}_{Y|X}$, $V(\mathsf{P}_X, \mathsf{P}_{Y|X})$ *is*

$$V(\mathsf{P}_X, \mathsf{P}_{Y|X}) = \sum_{x\in\mathcal{X}} \mathsf{P}_X(x) \left(\sum_{y\in\mathcal{Y}} \mathsf{P}_{Y|X}(y|x) \log^2 \frac{\mathsf{P}_{Y|X}(y|x)}{\mathsf{P}_Y(y)} - D(\mathsf{P}_{Y|X=x}\|\mathsf{P}_Y)^2 \right), \tag{53}$$

$V_1 = V(\mathsf{P}_U^*, \mathsf{P}_{Y|X} \circ \mathsf{P}_{X|U}^*)$ *and* $V_2 = V(\mathsf{P}_U^*, \mathsf{P}_{Z|X} \circ \mathsf{P}_{X|U}^*)$. $\mathsf{P}_{UX}^* = \mathsf{P}_U^* \mathsf{P}_{X|U}^*$ *is the probability distribution that achieve the secrecy capacity in Theorem 6.*

Before we go to the proof of theorem 7, we would first provide upper bound on secrecy constraint δ and error probability constraint ϵ.

Security.

We first provide a lemma based on privacy amplification below, which shows that there exists a hash function $g \in G$ such that the maximum total variation secrecy $S^{\max}(g(X)|\mathsf{P}_Z)$ is close to the average total variation secrecy $S^{\text{avg}}(g(X)|\mathsf{P}_Z)$.

Lemma 6. *For L satisfying equation (38) and $L \in \mathbb{N}$, let $\mathsf{P}_X = \mathsf{Q}_{\mathcal{X}}^{unif}$ be a uniform distribution over $\mathcal{X}$. G be a set of all functions satisfy the universal hash function criteria stated in subsection 2.1.2 and $|g^{-1}(m)| = L, \forall m \in \mathcal{M}$. Finally, let all $g \in G$ be uniformly distributed. Then, we have*

$$\mathsf{Pr}\left[\max_{m\in\mathcal{M}} d(\mathsf{P}_{Z|g^{-1}(m)}, \mathsf{P}_Z) \geq r + \mathbb{E}\left[d(\mathsf{P}_{Z|\mathcal{C}}, \mathsf{P}_Z)\right]\right] \leq |\mathcal{M}|\mathsf{e}^{-2Lr^2}, \tag{54}$$

where $r = \sqrt{\frac{\log(1+|\mathcal{M}|)}{2L}}$ and $\mathcal{C} = \{\bar{X}_1, ..., \bar{X}_L\}$ being a random codebook with codewords i.i.d. generated from P_X.

The proof can be founded in [10, Lemma. 3]. The main idea of the proof is using Hoeffding's reduction argument [34] to bound possibly dependent variables $X_1, ..., X_L$ by i.i.d. variables $\bar{X}_1, ..., \bar{X}_L$, and using the McDiarmid's inequality [35] (also called as bounded difference inequality). The lemma above guarantees that there exists a $g \in G$ satisfying

$$S^{\max}(M|Z) = \max_{m\in\mathcal{M}} d(\mathsf{P}_{Z|g^{-1}(m)}, \mathsf{P}_Z) \leq \mathbb{E}\left[d(\mathsf{P}_{Z|\mathcal{C},\mathsf{P}_Z})\right] + \sqrt{\frac{\log(1+|\mathcal{M}|)}{2L}}. \tag{55}$$

Finally, by Eq. (42) and for every Q_Z we can bound δ by

$$S^{\max}(M|Z) \leq \delta \leq \sup_{x\in\mathcal{C}} E_\gamma(\mathsf{P}_{Z|X=x}, \mathsf{Q}_Z) + \sqrt{\frac{\log(1+|\mathcal{M}|)}{2L}} + \sup_{x\in\mathcal{C}} \frac{1}{2}\sqrt{\frac{\gamma}{L}\mathbb{E}_{\mathsf{P}_{Z|X=x}}\left[\exp\left(-|\iota(x;Z) - \log\gamma|\right)\right]}. \tag{56}$$

Reliability.

After privacy amplification, the secrecy code is $M = g(X)$ with random encoder $\mathsf{P}_{X|M=m} = \mathsf{Q}^{\text{unif}}_{g^{-1}(m)}$. This means that, for each message $m \in \mathcal{M}$, the encoder randomly picks a codeword from the set $g^{-1}(m)$ and transmits it to the decoder. And then the decoder decodes the codeword $\hat{X}$ and outputs $\hat{M} = g(\hat{X})$. Since the message M is uniformly distributed by our construction, and furthermore, $|g^{-1}(\cdot)| = L$, for every $m \in \mathcal{M}$, we need L bits to identify the inverse of g function. Therefore, by random coding union bound and dependent testing bound, we can upper-bound the maximal error probability as

$$\mathsf{P}_{Y|M=m}\left[M \neq \hat{M}\right] \leq \epsilon \leq \min\{\epsilon_{DT,\max}(|\mathcal{M}|L-1), \epsilon_{RCU,\max}(|\mathcal{M}|L)\}. \tag{57}$$

With the above security and reiliabitlity proof, we are now ready to sketch the proof of theorem 7.

Proof sketch of Theorem 7

The first-order term is proved in section 3, so we would not repeat here. We turn our focus on the second-order terms.

First of all, the codes used for proving achievability of second-order terms is not i.i.d. codes, here we use constant composition codes [36, Sec. 4] instead. For each codeword $x^n(m)$ that composes a constant composition codes, it must satisfy

$$\frac{1}{n}\sum_{i=1}^{n} b(x_i(m)) \leq B, \tag{58}$$

where B is a constant and $b(x)$ is the cost function. The reason for using constant composition codes is that first of all, their E_γ are the same; secondly, they induce slightly better conditional variance V_1 and V_2. Also, in the method of types [37], for the codes in the same type class, they have the same cost. Now, the remaining of this problem is to prove the $Q^{-1}(\cdot)$ terms appeared in equation (52). Actually, we can observe that they are related to maximal total variation secrecy constraint δ and maximal error probability constraint ϵ. The basic idea is to apply Berry-Esseen theorem [38] to the E_γ terms in the bound of maximal error probability (Eq. (57)) and maximal total variation secrecy (Eq. (56)), as a result, we can get the $Q^{-1}(\cdot)$ terms. We can interpret them as the rate sacrificed for the error probability and secrecy. For more details, one may refer to [10, Thm. 13] and [36, Thm. 4.2].

4.2 Converse

The upper bound of average secrecy rate is given as below.

Theorem 8. *For a code $(n, |M|, \epsilon, \delta)^{avg}$ with $\epsilon + \delta < 1$, we have*

$$\mathrm{R}^{avg} \leq \max_{P_X}(I(X;Y|Z)) - \frac{V_c}{n}Q^{-1}(\epsilon+\delta) + \mathcal{O}(\frac{\log n}{n}), \tag{59}$$

where for P_X and $\mathsf{P}_{YZ|X}$, $V(\mathsf{P}_X, \mathsf{P}_{YZ|X})$ is

$$\tilde{V}(\mathsf{P}_X, \mathsf{P}_{YZ|X}) = \sum_{x\in\mathcal{X}} \mathsf{P}_X(x)\Big(\sum_{y,z} \mathsf{P}_{YZ|X}(y,z|x)\log^2 \frac{\mathsf{P}_{YZ|X}(y,z|x)}{\mathsf{P}_{Z|X}(z|x)\mathsf{P}_{Y|X}(y|x)} - D(\mathsf{P}_{YZ|X=x}\|\mathsf{P}_{Y|Z}\mathsf{P}_{Z|X=x})^2\Big), \tag{60}$$

and $V_c = \tilde{V}(\tilde{\mathsf{P}}_X^*, \mathsf{P}_{YZ|X})$, where $\tilde{\mathsf{P}}_X^*$ maximize the term $\max_{P_X}(I(X;Y|Z))$.

Before we proceed to the converse of the proof, we first state a proposition that give an converse bound on the rate $\mathrm{R}^{\mathrm{avg}}(n, \epsilon, \delta)$.

Proposition 9 (A numerically and analytically tractable converse bound [10])**.** *Let $\mathsf{Q}_{Y|Z} : \mathcal{Y} \to \mathcal{Z}$ be an arbitrary random transformation, and P_{XYZ} denote the distribution induced by the code. Then, every $(n, |\mathcal{M}|, \epsilon, \delta)^{\mathrm{avg}}$ secrecy code for the wiretap channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ satisfies*

$$|\mathcal{M}| \leq \inf_{\tau\in(0,1-\epsilon-\delta)} \frac{\tau+\delta}{\tau\beta_{1-\epsilon-\delta-\tau}(\mathsf{P}_{XYZ}, \mathsf{P}_{XZ}\mathsf{Q}_{Y|Z})}. \tag{61}$$

Proof sketch of Proposition 9

Proposition 9 [10, Thm. 12] can be proved based on the idea of meta-converse [10, 30, 39, 40]. The key observation is that the error probability and the secrecy index can be related through a suboptimal binary hypothesis testing between $\mathsf{P}_{M\hat{M}Z}$ and $\mathsf{Q}_{M\hat{M}Z}$:

$$T(m, \hat{m}, z) = 1\left\{m = \hat{m}, \mathsf{P}_{M|Z}(m|z) \leq \frac{1}{\eta|\mathcal{M}|}\right\}. \tag{62}$$

On the one hand, $\mathsf{Q}_{M\hat{M}Z}[T=1] \leq \frac{1}{|\mathcal{M}|\eta}$ is straightforward. On the other hand, by the reliability constraint $\mathsf{Pr}\left[M \neq \hat{M}\right] \leq \epsilon$, we have the lower bound for $\mathsf{P}_{M\hat{M}Z}[T=1]$ as follows:

$$\mathsf{P}_{M\hat{M}Z}[T=1] \geq 1-\epsilon-\mathsf{P}_{MZ}[P_{M|Z}(w|z) \geq \frac{1}{\eta|\mathcal{M}|}] \tag{63}$$

$$\overset{\text{(a)}}{\geq} 1-\epsilon-\frac{\delta}{1-\eta}, \tag{64}$$

where (a) follows from change of measure.

Finally, setting $\tau := \frac{\delta}{1-\eta-\delta}$ and using data-processing inequality of the Neyman-Pearson β function

$$\beta_\alpha(\mathsf{P}_{M\hat{M}Z}, \mathsf{Q}_{M\hat{M}Z}) \geq \beta_\alpha(\mathsf{P}_{M\hat{M}Z}, \mathsf{P}_{XZ}\mathsf{Q}_{Y|Z}), \tag{65}$$

we can attain the converse bound in (61).

Now we are ready for the proof sketch of Theorem 8.

Proof sketch of Theorem 8

As stated in the proof sketch of Theorem 7, the proof for Theorem 8 makes use of Berry-Esseen theorem [38] to the $\beta_{1-\epsilon-\delta-\tau}(\cdot)$ term to derive the $Q^{-1}(\cdot)$ term. We can interpret the $Q^{-1}(\cdot)$ as the necessary rate needed for average total variation secrecy constraint δ and average error probability constraint ϵ. For more details, please refer to [10, Thm. 13].

5 Wiretap Channel with Maximal Leakage

In this section, with the motivation given in Example 1, we consider the wiretap channel with maximal leakage (or Sibson mutual information of order infinity $\mathrm{I}^S_\infty(X;Z)$) as the security metric, which requires the following two conditions be simultaneously satisfied:

$$\begin{cases} \lim\limits_{n\to\infty} \mathsf{Pr}[M \neq \hat{M}] = 0; \\ \lim\limits_{n\to\infty} \mathcal{L}^{\max}_\infty(M_0 \to Z) = 0. \end{cases} \tag{66}$$

We first show that the stronger version of security-reliability constraint pair

$$\begin{cases} \lim\limits_{n\to\infty} \mathsf{Pr}[M \neq \hat{M}] = 0; \\ \lim\limits_{n\to\infty} \mathrm{D}_\infty(\mathsf{P}_{M_0Z^n} \| \mathsf{P}_{M_0}\mathsf{Q}_{Z^n}) = 0. \end{cases} \tag{67}$$

can be jointly satisfied by Theorem 10, and then reduce the result to the security metric of maximal leakage, which provides an inner bound for the capacity region.

Remark 3. The security constraint in (67) is stronger since

$$\mathrm{D}_\infty(\mathsf{P}_{M_0Z^n} \| \mathsf{P}_{M_0}\mathsf{Q}_{Z^n}) = \mathrm{I}^{\mathrm{S}}_\infty(M_0; Z^n) + \mathrm{D}_\infty(\mathsf{P}_{Z^n} \| \mathsf{Q}_{Z^n}) \tag{68}$$

$$\geq \mathrm{I}^{\mathrm{S}}_\infty(M_0; Z^n) = \mathcal{L}^{\max}_\infty(M_0 \to Z), \tag{69}$$

where the equality is due to the identity in [41, Eq. (48)].

By virtually adding a channel with memory $\mathsf{P}_{X^n|U^n}$ between the wiretap channel W and the encoder, Yu and Tan [14] completely characterized the following achievable region for the criterion (67).

Theorem 10 (Achievable region [14]). *Let $M \mapsto f_{\mathcal{C}}(M)$ be a random function induced by $\mathcal{C}$. For $s = (0,1] \cup \{\infty\}$, we have*

$$\mathscr{R}_{1+s}(\mathsf{Q}_Z) = \bigcup_{\tilde{\mathsf{P}}_{UX}:\tilde{\mathsf{P}}_X \in \mathcal{P}(\mathsf{P}_{Z|X},\mathsf{Q}_Z)} \left\{ (\mathrm{R}_0, \mathrm{R}_1) \middle| \begin{array}{l} \mathrm{R}_0 + \mathrm{R}_1 \le \mathrm{I}(U;Y)_{\tilde{\mathsf{P}}} \\ \mathrm{R}_0 \le \mathrm{I}(U;Y)_{\tilde{\mathsf{P}}} - \tilde{\mathrm{R}}_{1+s}\left(\tilde{\mathsf{P}}_{U|X}\tilde{\mathsf{P}}(X), \mathsf{P}_{Z|X}, \mathsf{Q}_Z\right) \end{array} \right\}, \tag{70}$$

where

$$\tilde{\mathrm{R}}_{1+s}\left(\tilde{\mathsf{P}}_{U|X}\tilde{\mathsf{P}}_X, \mathsf{P}_{Z|X}, \mathsf{Q}_Z\right) = \max_{\mathsf{P}_{Z|UX}} \left\{ -\frac{1+s}{s}\mathrm{D}\left(\tilde{\mathsf{P}}_{Z|UX} \middle\| \mathsf{P}_{Z|X}\tilde{\mathsf{P}}_{UX}\right) + \mathrm{D}\left(\tilde{\mathsf{P}}_{Z|U} \middle\| \mathsf{Q}_Z \middle| \tilde{\mathsf{P}}_U\right) \right\}. \tag{71}$$

A direct consequence of Theorem 10 is that under the "stronger" criterion (67), the secrecy capacity for $s = (0,1] \cup \{\infty\}$ is

$$\mathrm{C}_{1+s}(\mathsf{Q}_Z) := \max_{(\mathrm{R}_0,\mathrm{R}_1)\in\mathscr{R}_{1+s}(\mathsf{Q}_Z)} \mathrm{R}_0 = \max_{\tilde{\mathsf{P}}_{UX}:\tilde{\mathsf{P}}_X\in\mathcal{P}(\mathsf{P}_{Z|X},\mathsf{Q}_Z)} \mathrm{I}(U;Y) - \tilde{\mathrm{R}}_{1+s}\left(\tilde{\mathsf{P}}_{U|X}\tilde{\mathsf{P}}_X, \mathsf{P}_{Z|X}, \mathsf{Q}_Z\right). \tag{72}$$

Through the relation in Remark 3 and (72), the code achieving the capacity for the stronger criterion (67) can also satisfy the security requirement of maximal leakage, i.e. $\mathrm{C}_{1+s}(\mathsf{Q}_Z) \le \mathrm{C}_{\mathrm{MaxL}}$.

However, since there exist an extra term $\mathrm{D}_\infty(\mathsf{P}_{Z^n}\|\mathsf{Q}_{Z^n})$ in (68), we suppose that $\mathrm{C}_{\mathrm{MaxL}}$ can be strictly larger than $\mathrm{C}_{1+s}(\mathsf{Q}_Z)$. On the other hand, since Sibson Rényi mutual information is non-decreasing for α [41, Thm. 2], we have

$$\mathcal{L}_\infty^{\max}(M_0 \to Z) = \mathrm{I}_\infty^{\mathrm{S}}(M_0; Z^n) \ge \mathrm{I}_1^{\mathrm{S}}(M_0; Z^n) = \mathrm{I}(M_0; Z^n), \tag{73}$$

which indicates that $\mathrm{C}_{\mathrm{MaxL}}$ should not be greater than C_{S}, as defined in (43).

6 Discussion

Security metric	MI	TV	SS	MaxL
Secrecy capacity	C_{S}	C_{S}	C_{S}	$\mathrm{C}_{\mathrm{MaxL}}$

Table 1: Secrecy capacity of wiretap channels under security metrics: mutual information (MI), total variation secrecy (TV), semantic security (SS), and maximal leakage (MaxL).

The comparison of several security metrics is summarized in Table 1, where

$$\begin{aligned} \mathrm{C}_{\mathrm{S}} &= \max_{\mathsf{P}_{UX}} \mathrm{I}(U;Y) - \mathrm{I}(U;Z) \\ &\ge \mathrm{C}_{\mathrm{MaxL}} \ge \max_{\mathsf{P}_{UX}} \mathrm{I}(U;Y) - \tilde{\mathrm{R}}_\infty\left(\tilde{\mathsf{P}}_{U|X}\tilde{\mathsf{P}}_X, \mathsf{P}_{Z|X}, \mathsf{Q}_Z\right). \end{aligned}$$

Since maximal leakage is a stricter security metric compared to other metrics discussed in this report, C_{S} serves as an outer bound for $\mathrm{C}_{\mathrm{MaxL}}$ trivially.

Due to limited time, we have just found an inner bound for the secrecy capacity of the wiretap channel with maximal leakage in Section 5, and we are still trying to figure out at what transmission rate can the inner bound and outer bound meet. In other words, the full characterization of $\mathrm{C}_{\mathrm{MaxL}}$ has not been explored yet. Besides, the second-order analysis under maximal leakage constraint is still an open problem. Since channel resolvability has been used to characterize the secrecy exponent of Rényi divergence, we suspect that channel resolvability and privacy amplification are promising to be used to achieve an secrecy guarantee

in the second-order analysis of MaxL-WTC. Furthermore, it is intriguing to explore the trade-off between the maximal leakage rate and the error probability.

Recently in 2024, Lei Yu [15] has extended the solution to the Rényi resolvability problem from $\alpha \in [0, 2] \cup \{\infty\}$ to the entire range $\alpha \in \mathbb{R} \cup \{\pm\infty\}$. Hence, it would be interesting to investigate the wiretap channel in the 1-receiver 2-eavesdroppers setting under the Rényi resolvability security constraint for each eavesdropper.

Division of Work

	percentage	work items
Bo-Yu Yang	53%	Section 1, 2, 3, 4, 5, 6
Hsuan Yu	47%	Section 1, 2, 4, 6

References

[1] W. E. Mohammed Aziz Al Kabir and M. S. Sharif, "Securing iot devices against emerging security threats: Challenges and mitigation techniques," *Journal of Cyber Security Technology*, vol. 7, no. 4, pp. 199–223, 2023. [Online]. Available: https://doi.org/10.1080/23742917.2023.2228053

[2] J. Singh, T. Pasquier, J. Bacon, H. Ko, and D. Eyers, "Twenty security considerations for cloud-supported internet of things," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 269–284, 2016.

[3] C. E. Shannon, "Communication theory of secrecy systems," *The Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[4] A. D. Wyner, "The wire-tap channel," *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[5] M. Bloch, O. Günlü, A. Yener, F. Oggier, H. V. Poor, L. Sankar, and R. F. Schaefer, "An overview of information-theoretic security and privacy: Metrics, limits and applications," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 5–22, 2021.

[6] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[7] H. V. Poor, M. Goldenbaum, and W. Yang, "Fundamentals for iot networks: Secure and low-latency communications," in *Proceedings of the 20th International Conference on Distributed Computing and Networking*, 2019, pp. 362–364.

[8] C. Feng and H.-M. Wang, "Secure short-packet communications at the physical layer for 5g and beyond," *IEEE Communications Standards Magazine*, vol. 5, no. 3, pp. 96–102, 2021.

[9] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Advances in Cryptology – CRYPTO 2012*, R. Safavi-Naini and R. Canetti, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 294–311.

[10] W. Yang, R. F. Schaefer, and H. V. Poor, "Wiretap channels: Nonasymptotic fundamental limits," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4069–4093, 2019.

[11] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[12] R. Renner, "Security of quantum key distribution," *International Journal of Quantum Information*, vol. 6, no. 01, pp. 1–127, 2008.

[13] M. B. Parizi, E. Telatar, and N. Merhav, "Exact random coding secrecy exponents for the wiretap channel," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 509–531, 2016.

[14] L. Yu and V. Y. Tan, "Rényi resolvability and its applications to the wiretap channel," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1862–1897, 2018.

[15] L. Yu, "Rényi resolvability, noise stability, and anti-contractivity," *arXiv preprint arXiv:2402.07660*, 2024.

[16] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[17] M. H. Yassaee, M. R. Aref, and A. Gohari, "Non-asymptotic output statistics of random binning and its applications," in *2013 IEEE International Symposium on Information Theory*. IEEE, 2013, pp. 1849–1853.

[18] J. Liu, P. Cuff, and S. Verdú, "E_γ-resolvability," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2629–2658, 2016.

[19] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2000, pp. 351–368.

[20] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[21] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3047–3061, 2004.

[22] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journal of Computer and System Sciences*, vol. 28, no. 2, pp. 270–299, 1984. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022000084900709

[23] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1625–1657, 2019.

[24] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[25] H. Endo, M. Sasaki *et al.*, "Reliability and secrecy functions of the wiretap channel under cost constraint," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6819–6843, 2014.

[26] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM Journal on Computing*, vol. 17, no. 2, pp. 210–229, 1988. [Online]. Available: https://doi.org/10.1137/0217014

[27] W. Yang, R. F. Schaefer, and H. V. Poor, "Privacy amplification: Recent developments and applications," in *2018 International Symposium on Information Theory and Its Applications (ISITA)*, 2018, pp. 120–124.

[28] J. M. Renes, "On privacy amplification, lossy compression, and their duality to channel coding," *IEEE Transactions on Information Theory*, vol. 64, no. 12, pp. 7792–7801, 2018.

[29] J. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, 1979. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022000079900448

[30] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[31] A. Feinstein, "A new basic theorem of information theory," *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[32] I. Csiszár and J. Korner, "Broadcast channels with confidential messages," *IEEE transactions on information theory*, vol. 24, no. 3, pp. 339–348, 1978.

[33] M. Hayashi, H. Tyagi, and S. Watanabe, "Strong converse for a degraded wiretap channel via active hypothesis testing," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2014, pp. 148–151.

[34] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963. [Online]. Available: http://www.jstor.org/stable/2282952

[35] C. McDiarmid, *On the method of bounded differences*, ser. London Mathematical Society Lecture Note Series. Cambridge University Press, 1989, p. 148–188.

[36] V. Y. Tan *et al.*, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1-2, pp. 1–184, 2014.

[37] I. Csiszar, "The method of types [information theory]," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[38] N. Morton, "An introduction to probability and its applications," *American Journal of Human Genetics*, vol. 10, no. 1, p. 71, 1958.

[39] V. Y. Tan and M. R. Bloch, "Information spectrum approach to strong converse theorems for degraded wiretap channels," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1891–1904, 2015.

[40] G. Vazquez-Vilar, A. T. Campo, A. G. i Fàbregas, and A. Martinez, "Bayesian m-ary hypothesis testing: the meta-converse and verdú-han bounds are tight," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2324–2333, 2016.

[41] S. Verdú, "α-mutual information," in *2015 Information Theory and Applications Workshop (ITA)*. IEEE, 2015, pp. 1–6.