entropy

Article

Secrecy Capacity of the Extended Wiretap Channel II with Noise

Dan He [1], Wangmei Guo [1] and Yuan Luo [2,*]

[1] The State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China; dhe@stu.xidian.edu.cn (D.H.); wangmeiguo@mail.xidian.edu.cn (W.G.)
[2] Computer Science and Engineering Department, Shanghai Jiao Tong University, Shanghai 200240, China
* Correspondence: luoyuan@cs.sjtu.edu.cn; Tel.: +86-21-3420-5477

Academic Editors: Raúl Alcaraz Martínez and Kevin H. Knuth
Received: 28 June 2016; Accepted: 14 October 2016; Published: 31 October 2016

Abstract: The secrecy capacity of an extended communication model of wiretap channelII is determined. In this channel model, the source message is encoded into a digital sequence of length N and transmitted to the legitimate receiver through a discrete memoryless channel (DMC). There exists an eavesdropper who is able to observe arbitrary $\mu = N\alpha$ digital symbols from the transmitter through a second DMC, where $0 \leq \alpha \leq 1$ is a constant real number. A pair of an encoder and a decoder is designed to let the receiver be able to recover the source message with a vanishing decoding error probability and keep the eavesdropper ignorant of the message. This communication model includes a variety of wiretap channels as special cases. The coding scheme is based on that designed by Ozarow and Wyner for the classic wiretap channel II.

Keywords: wiretap channel; secrecy capacity; wiretap channel II; secrecy criteria

1. Introduction

The concept of the wiretap channel was first introduced by Wyner. In his celebrated paper [1], Wyner considered a communication model, where the transmitter communicated with the legitimate receiver through a discrete memoryless channel (DMC). Meanwhile, there existed an eavesdropper observing the digital sequence from the receiver through a second DMC. The goal was to design a pair of an encoder and a decoder such that the receiver was able to recover the source message perfectly, while the eavesdropper was ignorant of the message. That communication model is actually a degraded discrete memoryless wiretap channel.

After that, the communication models of wiretap channels have been studied from various aspects. Csiszár and Körner [2] considered a more general wiretap channel where the wiretap channel did not need to be a degraded version of the main channel, and common messages were also considered there. Other communication models of wiretap channels include wiretap channels with side information [3–8], compound wiretap channels [9–12] and arbitrarily-varying wiretap channels [13].

Ozarow and Wyner studied another kind of wiretap channel called the wiretap channel II [14]. The source message W was encoded into digital bits X^N and transmitted to the legitimate receiver via a binary noiseless channel. An eavesdropper could observe arbitrary $\mu = N\alpha$ digital bits from the receiver, where $0 < \alpha < 1$ is a constant real number not dependent on N.

Some extensions of the wiretap channel have been studied in recent years.

Cai and Yeung extended the wiretap channel II into the network scenario [15,16]. In that network model, the source message of length K was transmitted to the legitimate users through a network, and the eavesdropper was able to wiretap on at most $\mu < K$ edges. Cai and Yeung suggested using a linear "secret-sharing" method to provide security in the network. Instead of sending K message symbols, the source node sent μ random symbols and $K - \mu$ message symbols. Additionally, the code itself underwent a certain linear transformation. Cai and Yeung gave sufficient conditions for this transformation to guarantee security. They showed that as long as the field size is sufficiently large, a secure transformation existed. Some related work on wiretap networks was given in [17–20].

An extension of wiretap channel II was considered recently in [21,22]. The source message was encoded into the digital sequence X^N and transmitted to the legitimate receiver via a DMC. The eavesdropper could observe any $\mu = N\alpha$ digital bits of X^N from the transmitter. A pair of inner-outer bounds was given in [21], while the secrecy capacity with respect to the semantic secrecy criterion was established in [22]. The coding scheme in [21] was based on that developed by Ozarow and Wyner in [14], while the scheme in [22] was by Wyner's soft covering lemma. He et al. considered another extension of wiretap channel II in [23]. In that model, the eavesdropper observed arbitrary $\mu = N\alpha$ digital bits of the main channel output Y^N from the receiver. The capacity with respect to the strong secrecy criterion was established there. The proof of the coding theorem was based on Csiszár's almost independent coloring scheme. Some other work on wiretap channel II can be found in [24–26].

This paper considers a more general extension of wiretap channel II, where the eavesdropper observes arbitrary μ digital bits from the transmitter via a second DMC. The capacity with respect to the weak secrecy criteria is established. The coding scheme is based on that developed by Ozarow and Wyner in [14]. It is obvious that this communication model includes the general discrete memoryless wiretap channels, the wiretap channel II and the communication models discussed in [21–23] as special cases. Nevertheless, we should notice that the secrecy criteria considered in [22,23] are strictly stronger than those considered in this paper; see Figure 1.

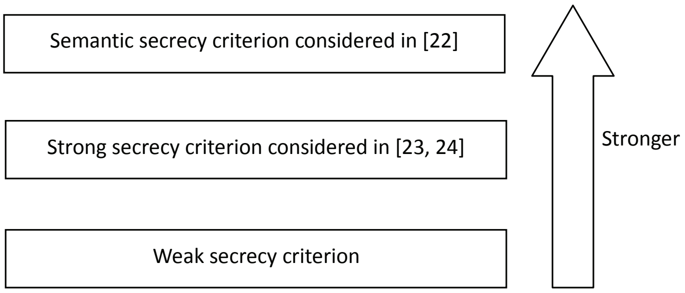

Semantic secrecy criterion considered in [22]
Strong secrecy criterion considered in [23, 24]
Weak secrecy criterion
Stronger

Figure 1. Comparison of different secrecy criteria.

The remainder of this paper is organized as follows. The formal statement and the main results are given in Section 2. The secrecy capacity is formulated in Theorem 1, whose proof is discussed in Section 4. Section 5 provides a binary example of this model. Section 6 gives a final conclusion of this paper.

2. Notation and Problem Statements

Throughout the paper, $\mathbb{N}$ is the set of positive integers and $[1:N] = \{1, 2, ..., N\}$ for any $N \in \mathbb{N}$. $\mathfrak{I}_\mu = \mathfrak{I}_\mu(N) = \{\mathcal{I} \subseteq [1:N] : |\mathcal{I}| = \mu\}$ represents the collection of subsets of $[1:N]$ with size μ.

Random variables, sample values and alphabets (sets) are denoted by capital letters, lower case letters and calligraphic letters, respectively. A similar convention is applied to random vectors and their sample values. For example, X^N represents a random N-vector $(X_1, X_2, ..., X_N)$, and x^N is a specific vector of X^N in $\mathcal{X}^N$. $\mathcal{X}^N$ is the N-th Cartesian power of $\mathcal{X}$.

Let '?' be a "dummy" letter. For any index set $\mathcal{I} \subseteq [1:N]$ and finite alphabet $\mathcal{X}$ not containing the "dummy" letter '?', denote:

$$\mathcal{X}_{\mathcal{I}}^N = \{(x_1, x_2, ..., x_N) : x_i \in \mathcal{X} \text{ if } i \in \mathcal{I}, \text{and } x_i = ? \text{ otherwise}\}.$$

For any given random vector $X^N = (X_1, X_2, ..., X_N)$ and index set $\mathcal{I} \subseteq [1:N]$,

- $X^N_{\mathcal{I}} = (X'_1, X'_2, ..., X'_N)$ is a "projection" of X^N onto $\mathcal{I}$ with $X'_n = X_n$ for $n \in \mathcal{I}$, and $X'_n =?$ otherwise.
- $X_{\mathcal{I}} = (X_i, i \in \mathcal{I})$ is a subvector of X^N.

The random vector $X^N_{\mathcal{I}}$ takes the value from $\mathcal{X}^N_{\mathcal{I}}$, while the random vector $X_{\mathcal{I}}$ takes the value from $\mathcal{X}^{|\mathcal{I}|}$.

Example 1. *Supposing that $N = 5$, the index set $\mathcal{I} = \{1, 3, 5\}$ and the random vector $X^N = (X_1, X_2, X_3, X_4, X_5)$, we have $X^N_{\mathcal{I}} = (X_1, ?, X_3, ?, X_5)$ and $X_{\mathcal{I}} = (X_1, X_3, X_5)$.*

The communication model in this paper, which is shown in Figure 2, is composed of an encoder, the main channel, the wiretap channel and a decoder. The definitions of these parts are from Definition 1 to Definition 4, respectively. The definition of achievability is in Definition 5.

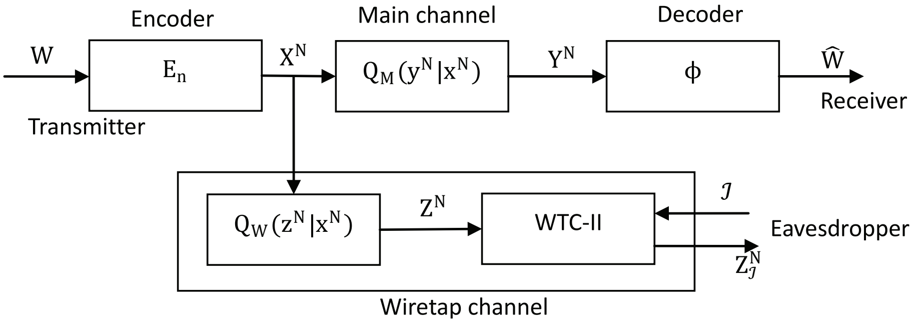


Figure 2. Communication model of wiretap channel II with noise.

Definition 1. *(Encoder) The source message W is uniformly distributed on the message set $\mathcal{W} = [1 : M]$. The (stochastic) encoder is specified by a matrix of conditional probability $En(x^N|w)$ for the channel input $x^N \in \mathcal{X}^N$ and message $w \in \mathcal{W}$.*

Definition 2. *(Main channel) The main channel is a DMC with finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$, where $? \notin \mathcal{X} \cup \mathcal{Y}$. The transition probability is $Q_M(y|x)$. Let X^N and Y^N denote the input and output of the main channel, respectively. It follows that for any $x^N \in \mathcal{X}^N$ and $y^N \in \mathcal{Y}^N$,*

$$\Pr\{X^N = x^N, Y^N = y^N\} = \Pr\{X^N = x^N\} Q^N_M(y^N|x^N)$$

where:

$$Q^N_M(y^N|x^N) = \prod_{n=1}^{N} Q_M(y_n|x_n).$$

Definition 3. *(Wiretap channel) The eavesdropper is able to observe arbitrary $\mu = N\alpha$ digital bits from the transmitter via another DMC, whose transition probability is denoted as $Q_W(z|x)$ with $x \in \mathcal{X}$ and $z \in \mathcal{Z}$. The alphabet $\mathcal{Z}$ does not contain the "dummy" letter '?' either. The input X^N and the output Z^N of the wiretap channel satisfy that:*

$$\Pr\{X^N = x^N, Z^N = z^N\} = \Pr\{X^N = x^N\} Q^N_W(z^N|x^N)$$

with:

$$Q^N_W(z^N|x^N) = \prod_{n=1}^{N} Q_W(z_n|x_n).$$

Supposing that the eavesdropper observes digital bits of the wiretap channel output Z^N whose indices lie in the observing index set $\mathcal{I}$, the subsequence obtained by the eavesdropper can then be denoted by $\tilde{Z}^N = Z^N_{\mathcal{I}}$. Therefore, the information on the source message (of each bit) exposed to the eavesdropper is denoted by:

$$\Delta = \max_{\mathcal{I}\in\mathfrak{I}_\mu} \frac{1}{N} I(W; Z^N_{\mathcal{I}}).$$

Remark 1. *Clearly, for given wiretap channel output $\tilde{Z}^N$, one can easily determine which subsequence of Z^N is observed by the eavesdropper. More precisely, $\tilde{Z}^N = Z^N_{\mathcal{I}}$ with $\mathcal{I} = \mathcal{I}(\tilde{Z}^N) = \{i : \tilde{Z}_i \neq ?\}$.*

Definition 4. *(Decoder) The decoder is a mapping $\phi : \mathcal{Y}^N \mapsto \mathcal{W}$, with Y^N as the input and $\hat{W} = \phi(Y^N)$ as the output. The average decoding error probability is defined as $P_e = \Pr\{W \neq \hat{W}\}$.*

Definition 5. *(Achievability) A non-negative real number R is said to be achievable, if for any $\epsilon > 0$, there exists an integer N_0, such that one can construct an (N, M) code satisfying:*

$$\frac{1}{N} \log M \geq R - \epsilon, \tag{1}$$

$$P_e \leq \epsilon \tag{2}$$

and:

$$\Delta < \epsilon, \tag{3}$$

where $N > N_0$. The capacity, or the maximal achievable transmission rate, of the communication model is denoted by C_s.

Remark 2. *Notice that the capacities defined in this paper are under the condition of negligible average decoding error probability, but one can construct the coding schemes for the negligible maximal decoding error probability through the standard techniques. See Appendix A for details.*

3. Main Result

Theorem 1. *The capacity of the communication model described in Figure 2 is:*

$$C_s = \max_{P_U\cdot P_{X|U}: U\to X\to YZ} [I(U;Y) - \alpha I(U;Z)],$$

where U is an auxiliary random variable distributed on $\mathcal{U}$ with $|\mathcal{U}| \leq |\mathcal{X}|$.

The converse half of Theorem 1 can be established quite similarly to the method of establishing the converse of Theorem 2 in [22], and hence, we omit it here. The direct part of Theorem 1 is given in Section 4.

Corollary 1. *When $\alpha = 1$, the communication model is transformed into a general discrete memoryless wiretap channel, whose capacity is formulated by:*

$$C_s = \max_{P_U\cdot P_{X|U}: U\to X\to YZ} [I(U;Y) - I(U;Z)].$$

The result coincides with that of Corollary 2 in [2]. In particular, if $X \to Y \to Z$ forms a Markov chain, the capacity is further deduced by:

$$\begin{aligned} C_s &= \max_{P_U \cdot P_{X|U}: U \to X \to Y \to Z} [I(U;Y) - I(U;Z)] \\ &\overset{(a)}{=} \max_{P_U \cdot P_{X|U}: U \to X \to Y \to Z} [I(U;Y|Z)] \\ &\overset{(b)}{=} \max_{P_X: X \to Y \to Z} [I(X;Y|Z)] \\ &\overset{(c)}{=} \max_{P_X: X \to Y \to Z} [I(X;Y) - I(X;Z)], \end{aligned}$$

where (a) follows because $U \to Y \to Z$ forms a Markov chain, (b) follows because the Markov chain $U \to X \to Y$ for a given Z implies that $I(U;Y|Z) \leq I(X;Y|Z)$ and the equality holds if and only if $U = X$ and (c) follows because $X \to Y \to Z$ forms a Markov chain.

Corollary 2. *When $Y = Z$, i.e., the eavesdropper observes $\mu = N\alpha$ digital bits from the receiver, the communication model is transformed into that studied in [23]. In this case, the capacity is formulated by:*

$$C_s = \max_{P_U \cdot P_{X|U}: U \to X \to Y} (1-\alpha) I(U;Y) = \max_{P_X} (1-\alpha) I(X;Y),$$

where the last equality follows because the Markov chain $U \to X \to Y$ implies that $I(U;Y) \leq I(X;Y)$ and the equality holds if and only if $U = X$.

Notice that [23] considered the capacity with respect to the strong secrecy criterion, while the current paper considers that with the weak secrecy criterion. Therefore, Theorem 1 in [23] and Corollary 2 in the current paper indicate that the capacity with respect to the strong secrecy criterion is identical to that with the weak secrecy criterion.

Corollary 3. *When $Z = X$, i.e., the eavesdropper observes $\mu = N\alpha$ digital bits from the transmitter, the communication model is transformed into that studied in [21,22]. In this case, the capacity is formulated by:*

$$C_s = \max_{P_U \cdot P_{X|U}: U \to X \to Y} [I(U;Y) - \alpha I(U;X)].$$

The channel model described in Corollary 3 was first studied in [21], and a pair of inner and outer bounds was given there. The secrecy capacity with respect to the semantic secrecy criterion, which is identical to the capacity with the weak criterion given in Corollary 3, was established in [22].

4. Direct Part of Theorem 1

This section proves the direct part of Theorem 1. Let an arbitrary quadruple of random variables (U^*, X^*, Y^*, Z^*) be given, which satisfy that:

$$\Pr\{U^* = u, X^* = x, Y^* = y, Z^* = z\} = P_{U^*}(u) P_{X^*|U^*}(x|u) Q_M(y|x) Q_W(z|x) \tag{4}$$

for $(u, x, y, z) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. The goal of this section is to establish that every real number R satisfying $0 \leq R < I(U^*;Y^*) - \alpha I(U^*;Z^*)$ is achievable. In fact, it suffices to prove the achievability of every R satisfying:

$$0 \leq R < I(X^*;Y^*) - \alpha I(X^*;Z^*). \tag{5}$$

To show this, suppose that every transmission rate R satisfying Equation (5) is achievable. For any random variable U^* satisfying Equation (4), the encoder could deliberately increase the noise of the communication system by inserting a virtual noisy channel Q_V at the transmitting port of the system, such that:

$$Q_V(x|u) = P_{X^*|U^*}(x|u).$$

This would create the virtual communication system depicted in Figure 3, where the transition matrix of the main channel is:

$$\tilde{Q}_M(y|u) = P_{Y^*|U^*}(y|u) = \sum_{x\in\mathcal{X}} Q_V(x|u)Q_M(y|x)$$

and the transition matrix of the wiretap channel is:

$$\tilde{Q}_W(z|u) = P_{Z^*|U^*}(y|u) = \sum_{x\in\mathcal{X}} Q_V(x|u)Q_W(y|x).$$

It is clear that every real number $0 < R \leq I(U^*;Y^*) - \alpha I(U^*;Z^*)$ is achievable for the virtual system and hence is achievable for the original system.

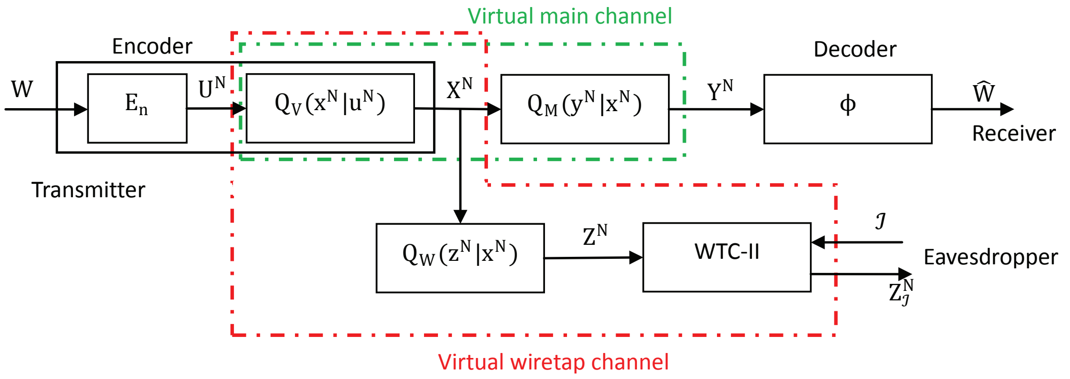


Figure 3. Adding a virtual channel Q_V to the communication system.

In the remainder of this section, it will be shown that every transmission rate R satisfying Equation (5) is achievable. To be precise, for any ϵ and τ satisfying $0 < \epsilon \leq \tau < I(X^*;Y^*) - \alpha I(X^*;Z^*)$, we need to establish that there exists an (N, M) code, such that:

$$\frac{1}{N}\log M \geq I(X^*;Y^*) - \alpha I(X^*;Z^*) - \tau > 0, \tag{6}$$

$$\Delta = \max_{\mathcal{I}\in\mathfrak{I}_\mu} \frac{1}{N} I(W;Z^N_{\mathcal{I}}) < \epsilon, \tag{7}$$

and:

$$P_e < \epsilon \tag{8}$$

when N is sufficiently large.

The coding scheme is based on the scheme developed by Ozarow and Wyner [14] for the classic wiretap channel II. In that channel model, for each wiretap channel output z^N, there exists a collection of codewords that are "consistent" with it, namely the codewords that could produce the wiretap channel output z^N for some observing index $\mathcal{I}$. Ozarow and Wyner constructed a secure partition, such that the number of "consistent" codewords, for every wiretap channel output z^N, in each sub-code is less than a constant integer. However, it is not feasible to consider the "consistent" codewords in our model, where the wiretap channel may be noisy. Instead, we construct a secure partition such that the number of codewords, jointly typical with the wiretap channel output z^N, in each sub-code is less than a constant integer.

The proof is organized as follows. Firstly, Section 4.1 gives some definitions on the typicality of μ-subsequences and lists some basic results. Then, the construction of the encoder and decoder is introduced in Section 4.2. The key point is to generate a "good" codebook with the desired partition to ensure secrecy. Thirdly, as the main part of the proof, Section 4.3 shows the existence of a "good" codebook with the desired partition. For any $\epsilon > 0$ and $\tau > 0$, the proof that the coding scheme in Section 4.2 satisfies the requirements of the transmission rate, reliability and security, namely Formulas (6) to (8), is finally detailed in Section 4.4.

4.1. Typicality

The definitions of letter typicality on a given index set follow from those briefly introduced in [23]. We list them in this subsection for the sake of completeness.

Firstly, the original definitions of letter typicality are given as follows. Please refer to Chapter 1 in [27] for more details.

For any $\delta \geq 0$, the δ-letter typical set $T_\delta^N(P_X)$ with respect to the probability distribution P_X on $\mathcal{X}$ is the set of $x^N \in \mathcal{X}^N$ satisfying:

$$|\frac{1}{N}\mathbf{N}(a|x^N) - P_X(a)| \leq \delta P_X(a) \text{ for all } a \in \mathcal{X},$$

where $\mathbf{N}(a|x^N)$ is the number of positions of x^N having the letter $a \in \mathcal{X}$.

Similarly, let $\mathbf{N}(a,b|x^N,y^N)$ be the number of times that the pair (a,b) occurs in the sequence of pairs $(x_1,y_1),(x_2,y_2),...,(x_N,y_N)$. The jointly typical set $T_\delta^N(P_{XY})$, with respect to the joint probability distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$, is the set of sequence pairs $(x^N,y^N) \in \mathcal{X}^N \times \mathcal{Y}^N$ satisfying:

$$|\frac{1}{N}\mathbf{N}(a,b|x^N,y^N) - P_{XY}(a,b)| \leq \delta P_{XY}(a,b)$$

for all $(a,b) \in \mathcal{X} \times \mathcal{Y}$.

For any given $x^N \in \mathcal{X}^N$, the conditionally typical set of x^N with respect to the joint mass function P_{XY} is defined as:

$$T_\delta^N(P_{XY}|x^N) = \{y^N \in \mathcal{Y}^N : (x^N,y^N) \in T_\delta^N(P_{XY})\}.$$

The definitions on the typicality of μ-subsequences on index set $\mathcal{I} \in \mathfrak{I}_\mu$ and some basic results are given as follows.

Definition 6. *Given a random variable X on $\mathcal{X}$, the letter typical set $\tilde{T}_\mathcal{I}^N[X]_\delta$ with respect to X on the index set $\mathcal{I} \in \mathfrak{I}_\mu$ is the set of $x^N \in \mathcal{X}_\mathcal{I}^N$ such that $x_\mathcal{I} \in T_\delta^\mu(P_X)$, where $x_\mathcal{I}$ is the μ-subvector of x^N and P_X is the probability mass function of the random variable X.*

Example 2. *Let X be a random variable on the binary set $\mathcal{X} = \{0,1\}$, such that $P_X(0) = 1 - P_X(1) = \frac{1}{5}$. Set $N = 10$, $\delta = 0.1$ and $\mathcal{I} = [1:5]$. Then:*

- *the sequence $x^N = 0111100000$ is out of $T_\delta^N(P_X)$ while $x_\mathcal{I}^N = 01111?????$ belongs to $\tilde{T}_\mathcal{I}^N[X]_\delta$;*
- *the sequence $x'^N = 1111101110$ belongs to $T_\delta^N(P_X)$ while $x_\mathcal{I}'^N = 11111?????$ is out of $\tilde{T}_\mathcal{I}^N[X]_\delta$;*
- *the sequence $x''^N = 0111111110$ belongs to $T_\delta^N(P_X)$, and $x_\mathcal{I}''^N = 01111?????$ belongs to $\tilde{T}_\mathcal{I}^N[X]_\delta$.*

Remark 3. *(Theorem 1.1 in [27]) Suppose that $X_1, X_2, ..., X_N$ are N i.i.d. random variables with the same generic probability distribution as that of X. For any given $\mathcal{I} \in \mathfrak{I}_\mu$ and $\delta < m_X$,*

1. *if $x^N \in \tilde{T}_\mathcal{I}^N[X]_\delta$, then*

$$2^{-\mu(1+\delta)H(X)} \leq \Pr\{X_\mathcal{I}^N = x^N\} \leq 2^{-\mu(1-\delta)H(X)},$$

2. $\Pr\{X_\mathcal{I}^N \in \tilde{T}_\mathcal{I}^N[X]_\delta\} > 1 - \tilde{\epsilon}_1$,

where:

$$\tilde{\epsilon}_1 = \tilde{\epsilon}_1(\mu, \delta, m_X) = 2|\mathcal{X}|e^{-\mu\delta^2 m_X}$$

and $m_X = \min_{x\in\mathcal{X}:P_X(x)>0} P_X(x)$.

Definition 7. *Let* (X,Y) *be a pair of random variables with the joint probability mass function* P_{XY} *on* $\mathcal{X}\times\mathcal{Y}$. *The jointly typical set* $\tilde{T}^N_{\mathcal{I}}[XY]_\delta$ *with respect to* (X,Y) *on the index set* $\mathcal{I}\in\mathfrak{I}_\mu$ *is the set of* $(x^N,y^N)\in\mathcal{X}^N_{\mathcal{I}}\times\mathcal{Y}^N_{\mathcal{I}}$ *satisfying* $(x_{\mathcal{I}},y_{\mathcal{I}})\in T^\mu_\delta(P_{XY})$, *where* $x_{\mathcal{I}}$ *and* $y_{\mathcal{I}}$ *are the subvectors of* x^N *and* y^N, *respectively.*

Definition 8. *For any given* $x^N\in\tilde{T}^N_{\mathcal{I}}[X]_\delta$ *with* $\mathcal{I}\in\mathfrak{I}_\mu$, *the conditionally-typical set of* x^N *on the index set* $\mathcal{I}$ *is defined as:*

$$\tilde{T}^N_{\mathcal{I}}[XY|x^N]_\delta = \{y^N : (x^N,y^N)\in\tilde{T}^N_{\mathcal{I}}[XY]_\delta\}.$$

Remark 4. *Let* (X^N,Y^N) *be a pair of random sequences with the conditional mass function:*

$$\Pr\{Y^N = y^N|X^N = x^N\} = \prod_{i=1}^{N} P_{Y|X}(y_i|x_i) \tag{9}$$

for $x^N\in\mathcal{X}^N$ *and* $y^N\in\mathcal{Y}^N$. *Then, for any index set* $\mathcal{I}\in\mathfrak{I}_\mu$, $x^N\in\tilde{T}^N_{\mathcal{I}}[X]_\delta$ *and* $y^N\in\tilde{T}^N_{\mathcal{I}}[XY|x^N]_\delta$, *it follows that:*

$$\begin{aligned}2^{-\mu(1+\delta)H(Y|X)} &\le \Pr\{Y^N_{\mathcal{I}} = y^N|X^N_{\mathcal{I}} = x^N\}\\ &\le 2^{-\mu(1-\delta)H(Y|X)}.\end{aligned}$$

Corollary 4. *For any* $\mathcal{I}\in\mathfrak{I}_\mu$ *and* $x^N\in\tilde{T}^N_{\mathcal{I}}[X]_\delta$, *it follows that* $|\tilde{T}^N_{\mathcal{I}}[XY|x^N]_\delta| < 2^{N\alpha(1+\delta)H(Y|X)}$.

Corollary 5. *Let* $Y_1, Y_2, ..., Y_N$ *be* N *i.i.d. random variables with the same probability distribution as that of* Y. *For any* $\mathcal{I}\in\mathfrak{I}_\mu$ *and* $x^N\in\tilde{T}^N_{\mathcal{I}}[X]_\delta$, *it follows that:*

$$\Pr\{Y^N\in\tilde{T}^N_{\mathcal{I}}[XY|x^N]_\delta\} < 2^{-N[\alpha I(X;Y)-2\delta H(Y)]}.$$

Remark 5. *(Theorem 1.2 in [27]) Let* (X^N,Y^N) *be a pair of random sequences satisfying* (9). *For any index set* $\mathcal{I}\in\mathfrak{I}_\mu$, $0\le\delta<m_{XY}$ *and* $x^N\in\tilde{T}^N_{\mathcal{I}}[X]_\delta$, *it is satisfied that:*

$$\Pr\{Y^N_{\mathcal{I}}\in\tilde{T}^N_{\mathcal{I}}[XY|x^N]_{2\delta}|X^N_{\mathcal{I}} = x^N\} > 1-\tilde{\epsilon}_2,$$

where:

$$\tilde{\epsilon}_2 = \tilde{\epsilon}_2(\mu,\delta,m_{XY}) = 2|\mathcal{X}||\mathcal{Y}|e^{-\mu m_{XY}\frac{\delta^2}{1+2\delta}}$$

and $m_{XY} = \min_{(x,y)\in\mathcal{X}\times\mathcal{Y}:P_{XY}(x,y)>0} P_{XY}(x,y)$.

4.2. Code Construction

Suppose that the triple of random variables (X^*,Y^*,Z^*) is given and fixed.

Codeword generation: The random codebook $\mathbf{C} = \{X^{*N}(l)\}_{l=1}^{M'}$ is an ordered set of M' i.i.d. random vectors with mass function $\Pr\{X^{*N}(l) = x^N\} = \prod_{i=1}^{N} P_{X^*}(x_i)$, where:

$$M' = 2^{N[I(X^*;Y^*)-\tau-\tau_d]} \tag{10}$$

for some $\tau_d > 0$.

Codeword partition: Given a specific sample value $\mathcal{C} = \{x^N(l)\}_{l=1}^{M'}$ of M' randomly-generated codewords, let W' be a random variable uniformly distributed on $[1:M']$ and $X^N(\mathcal{C}) = x^N(W')$ be the random sequence uniformly distributed on $\mathcal{C}$. Set $R = I(X^*;Y^*) - \alpha I(X^*;Z^*) - \tau$, and partition $\mathcal{C}$ into:

$$M = 2^{NR} = 2^{N[I(X^*;Y^*)-\alpha I(X^*;Z^*)-\tau]} \tag{11}$$

subsets $\{\mathcal{C}_m\}_{m=1}^{M}$ with the same cardinality. Let $\tilde{W}$ be the index of sub-code containing $X^N(\mathcal{C})$, i.e., $X^N(\mathcal{C}) \in \mathcal{C}_{\tilde{W}}$. We need to find a partition of the codebook $\mathcal{C}$ satisfying that:

$$\max_{\mathcal{I}\in\mathfrak{I}_\mu} \frac{1}{N} I(\tilde{W}; Z^N_{\mathcal{I}}(\mathcal{C})) < \epsilon, \tag{12}$$

where $Z^N(\mathcal{C})$ is the output of the wiretap channel when taking $X^N(\mathcal{C})$ as the input, i.e.,

$$\Pr\{X^N(\mathcal{C}) = x^N, Z^N(\mathcal{C}) = z^N\} = \Pr\{X^N(\mathcal{C}) = x^N\} \prod_{i=1}^{N} Q_W(z_i|x_i).$$

for $x^N \in \mathcal{X}^N$ and $z^N \in \mathcal{Z}^N$. If there is no such desired partition, declare an encoding error.

Remark 6. *We call the codebook $\mathcal{C}$ an ordered set because each codeword in the codebook is treated as unique, even if its value may be the same as the other codewords.*

Encoder: Suppose that a desired partition $\{\mathcal{C}_m\}_{m=1}^{M}$ on a specific codebook $\mathcal{C}$ is given. When the source message W is to be transmitted, the encoder uniformly randomly chooses a codeword from the sub-code $\mathcal{C}_W$ and transmits it to the main channel.

In this encoding scheme, each message is related to a unique sub-code, which is sometimes called a bin. Therefore, we would call this kind of coding scheme the random binning scheme.

Remark 7. *For a given codebook $\mathcal{C}$ and a desired partition applied to the encoder, let X^N and Z^N be the input and output of the wiretap channel respectively, when the source message W is transmitted. It is clear that (W, X^N, Z^N) and $(\tilde{W}, X^N(\mathcal{C}), Z^N(\mathcal{C}))$ share the same joint distribution.*

Decoder: Supposing that the output of the main channel is y^N. The decoder tries to find a unique sequence $x^N(\hat{w}, \hat{j})$, such that $(x^N(\hat{w}, \hat{j}), y^N) \in T^N_\delta(P_{X^*Y^*})$ and decodes $\hat{w}$ as the estimation of the transmitted source message. If there is none or there is more than one satisfied $x^N(\hat{w}, \hat{\imath})$, the encoder chooses a constant w_0 as $\hat{w}$.

4.3. Proof of the Existence of a "Good" Codebook with a Secure Partition

This subsection proves the existence of a class of "good" codebooks, on which there exist secure partitions, such that Equation (12) holds, when N is sufficiently large and δ is sufficiently small. Moreover, those kinds of "good" codebooks can be randomly generated with probability → 1 as $N \to \infty$. The notation in Section 4.2 will continue to be used in this subsection.

A formal definition of "good" codebooks is given by the following.

Definition 9. *A codebook $\mathcal{C}$ is called "good" if it is satisfied that:*

$$|\tilde{T}(\mathcal{C}, \mathcal{I})| > (1 - 2\epsilon_1)M' \text{ for all } \mathcal{I} \in \mathfrak{I}_\mu \tag{13}$$

and:

$$|\tilde{T}(\mathcal{C}, z^N, \mathcal{I})| < 2^{N(I(X^*;Y^*)-\alpha I(X^*;Z^*)-\tau-\frac{\tau_d}{2})} \text{ for all } \mathcal{I} \in \mathfrak{I}_\mu \text{ and } z^N \in \tilde{T}^N_{\mathcal{I}}[Z^*]_{2\delta}, \tag{14}$$

where:

$$\tilde{T}(\mathcal{C}, \mathcal{I}) = \{x^N \in \mathcal{C} : x^N_{\mathcal{I}} \in \tilde{T}^N_{\mathcal{I}}[X^*]_\delta\}$$

is the set of typical codewords on the index set $\mathcal{I}$*,*

$$\tilde{T}(\mathcal{C}, z^N, \mathcal{I}) = \{x^N \in \mathcal{C} : x^N_{\mathcal{I}} \in \tilde{T}^N_{\mathcal{I}}[X^*Y^*|z^N]_{2\delta}\}$$

is the set of codewords jointly typical with z^N *on the index set* $\mathcal{I}$ *and:*

$$\epsilon_1 = \epsilon_1(\mu, \delta, m_{X^*}) = 2|\mathcal{X}|e^{-\mu\delta^2 m_{X^*}}. \tag{15}$$

The main results of this subsection are summarized as the following three lemmas. Lemma 1 claims the existence of "good" codebooks; Lemma 2 constructs a special class of partitions on the "good" codebooks; and Lemma 3 proves that the partitions constructed by Lemma 2 are secure.

Lemma 1. *Let* $\mathbf{C}$ *be the random codebook generated by the scheme introduced in Section 4.2. If* $\delta < m_{X^*}$*, the probability of* $\mathbf{C}$ *being "good" is bounded by:*

$$\Pr\{\mathbf{C} \text{ is "good"}\} > 1 - \epsilon_3 - \epsilon_4,$$

where:

$$\epsilon_3 = \exp_2[N - (2 - \log e)\epsilon_1 M'], \tag{16}$$

$$\epsilon_4 = \exp_2[2N + (2^{-N(\tau_d - 4\delta H(X^*))} \log e - 2^{-\frac{1}{2}N\tau_d})M], \tag{17}$$

and ϵ_1 *is given by Equation (15).*

The proof of Lemma 1 is detailed in Appendix B.

Remark 8. *It can be verified that* $\epsilon_3, \epsilon_4 \to 0$ *as* $N \to \infty$*, if* δ *is sufficiently small. Therefore, one can obtain a "good" codebook with probability* $\to 1$*.*

Lemma 2. *For any given codebook* $\mathcal{C}$ *satisfying Equation (14) with* $M' = 2^{N(I(X^*;Y^*) - \tau - \tau_d)}$ *codewords, there exists a secure equipartition* $\{\mathcal{C}_m\}_{m=1}^M$ *on it, such that:*

$$|\tilde{T}(\mathcal{C}, \mathcal{I}, z^N) \cap \mathcal{C}_m| < L \tag{18}$$

for all $1 \le m \le M, \mathcal{I} \in \mathfrak{I}_\mu$ *and* $z^N \in T^N_{\mathcal{I}}[Y^*]_{2\delta}$*, if* $L > \frac{2(R+5)}{\tau_d}$*, where:*

$$R = \frac{1}{N}\log M = I(X^*; Y^*) - \alpha I(X^*; Z^*) - \tau.$$

The proof of Lemma 2 is discussed in Appendix C.

Lemma 3. *For any* $0 < \delta < m_{X^*Y^*}$ *and secure partition* $\{\mathcal{C}_m\}_{m=1}^M$ *on a "good" codebook* $\mathcal{C}$*, it follows that:*

$$\max_{\mathcal{I} \in \mathfrak{I}_\mu} \frac{1}{N} I(\tilde{W}; Z^N_{\mathcal{I}}(\mathcal{C})) < \tau_d + \frac{2 + \log L}{N} + (4\delta + 2\epsilon_5)\log|\mathcal{Z}| + \epsilon_5 \log|\mathcal{X}|, \tag{19}$$

where:

$$\epsilon_5 = 2\epsilon_1 + \epsilon_2 \tag{20}$$

and:

$$\epsilon_2 = \epsilon_2(\mu, \delta, m_{X^*Z^*}) = 2|\mathcal{X}||\mathcal{Z}|e^{-\mu m_{X^*Z^*}\frac{\delta^2}{1+2\delta}}.$$

Remark 9. *Formula (12) is finally established from the fact that the right-hand side of Equation (19) converges to zero as* $\tau_d \to 0$ *and* $N \to \infty$*.*

Proof of Lemma 3. By a way similar to establishing Equation (22) in [2], for every $\mathcal{I} \in \mathfrak{I}_\mu$, it follows that:

$$\begin{aligned}
H(\tilde{W}|Z_{\mathcal{I}}^N(\mathcal{C})) &= H(\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C})) - H(Z_{\mathcal{I}}^N(\mathcal{C})) \\
&= H(\tilde{W}, X^N(\mathcal{C}), Z_{\mathcal{I}}^N(\mathcal{C})) - H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C})) - H(Z_{\mathcal{I}}^N(\mathcal{C})) \\
&= H(\tilde{W}, X^N(\mathcal{C})) + H(Z_{\mathcal{I}}^N(\mathcal{C})|\tilde{W}, X^N(\mathcal{C})) - H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C})) - H(Z_{\mathcal{I}}^N(\mathcal{C})) \\
&= H(\tilde{W}, X^N(\mathcal{C})) + H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C})) - H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C})) - H(Z_{\mathcal{I}}^N(\mathcal{C})),
\end{aligned}$$

where the last equality follows because $\tilde{W} \to X^N(\mathcal{C}) \to Z_{\mathcal{I}}^N(\mathcal{C})$ forms a Markov chain. Therefore:

$$\begin{aligned}
I(\tilde{W}; Z_{\mathcal{I}}^N(\mathcal{C})) &= H(\tilde{W}) - H(\tilde{W}|Z_{\mathcal{I}}^N(\mathcal{C})) \\
&= H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C})) + H(Z_{\mathcal{I}}^N(\mathcal{C})) - H(X^N(\mathcal{C})|\tilde{W}) - H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C})).
\end{aligned} \tag{21}$$

The terms in the rightmost side of Equation (21) are bounded as follows.

- Upper bound of $H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}))$. On account of the fact that $X^N(\mathcal{C})$ is uniformly distributed on the "good" codebook $\mathcal{C}$ (cf. Equation (13)), it follows that:

$$\Pr\{X_{\mathcal{I}}^N(\mathcal{C}) \in \tilde{T}_{\mathcal{I}}^N[X^*]_\delta\} > 1 - 2\epsilon_1$$

for every $\mathcal{I} \in \mathfrak{I}_\mu$. Combining Remark 5 yields:

$$\Pr\{(X_{\mathcal{I}}^N(\mathcal{C}), Z_{\mathcal{I}}^N(\mathcal{C})) \in \tilde{T}_{\mathcal{I}}^N[X^*Z^*]_{2\delta}\} > 1 - \epsilon_5$$

for every $\mathcal{I} \in \mathfrak{I}_\mu$, where ϵ_5 is given by Equation (20). Denote:

$$U_{\mathcal{I}} = \begin{cases} 0 & \text{if } (X_{\mathcal{I}}^N(\mathcal{C}), Z_{\mathcal{I}}^N(\mathcal{C})) \in \tilde{T}_{\mathcal{I}}^N[X^*Z^*]_{2\delta}, \\ 1 & \text{otherwise.} \end{cases}$$

It follows that:

$$\Pr\{U_{\mathcal{I}} = 1\} < \epsilon_5 \tag{22}$$

for every $\mathcal{I} \in \mathfrak{I}_\mu$. Moreover, on account of the property of (18), we also have:

$$H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}), U_{\mathcal{I}} = 0) \le logL. \tag{23}$$

Therefore:

$$\begin{aligned}
& H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C})) \\
\le\ & H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}), U_{\mathcal{I}}) + H(U_{\mathcal{I}}) \\
\le\ & H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}), U_{\mathcal{I}} = 0) + H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}), U_{\mathcal{I}} = 1) \Pr\{U_{\mathcal{I}} = 1\} + H(U_{\mathcal{I}}) \\
\overset{(a)}{\le}\ & 1 + \log L + H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}), U_{\mathcal{I}} = 1) \Pr\{U_{\mathcal{I}} = 1\} \\
\overset{(b)}{\le}\ & 1 + \log L + N\epsilon_5 \log |\mathcal{X}|,
\end{aligned} \tag{24}$$

where (a) follows from Equation (23) and the fact that $U_{\mathcal{I}}$ is binary and (b) follows from Equation (22).

- The value of $H(Z_{\mathcal{I}}^N(\mathcal{C}))$ is upper bounded as:

$$\begin{array}{rl} & H(Z_{\mathcal{I}}^N(\mathcal{C})) \\ \leq & H(Z_{\mathcal{I}}^N(\mathcal{C})|U_{\mathcal{I}}) + H(U_{\mathcal{I}}) \\ \leq & H(Z_{\mathcal{I}}^N(\mathcal{C})|U_{\mathcal{I}} = 0) + H(Z_{\mathcal{I}}^N(\mathcal{C})|U_{\mathcal{I}} = 1)\Pr\{U_{\mathcal{I}} = 1\} + H(U_{\mathcal{I}}) \\ \overset{(a)}{\leq} & 1 + N\alpha(1+2\delta)H(Z^*) + H(X^N(\mathcal{C})|\tilde{W}, Z_{\mathcal{I}}^N(\mathcal{C}), U_{\mathcal{I}} = 1)\Pr\{U_{\mathcal{I}} = 1\} \\ \overset{(b)}{\leq} & 1 + N\alpha(1+2\delta)H(Z^*) + N\epsilon_5 \log|\mathcal{Z}|, \end{array} \tag{25}$$

  where (a) follows the facts that $U_{\mathcal{I}}$ is binary and $Z_{\mathcal{I}}^N(\mathcal{C}) \in \tilde{T}_{\mathcal{I}}^N[Z^*]_{2\delta}$ when $U_{\mathcal{I}} = 0$ and (b) follows from Equation (22).
- Recalling that when given $\tilde{W} = w$, the random vector $X^N(\mathcal{I})$ is uniformly distributed on $\mathcal{C}_w$, we have:

$$H(X^N(\mathcal{C})|\tilde{W}) = \log\frac{M'}{M} = N(\alpha I(X^*;Z^*) - \tau_d). \tag{26}$$

- Lower bound of $H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C}))$. For any x^N satisfying that $x_{\mathcal{I}}^N \in \tilde{T}_{\mathcal{I}}^N[X^*]_{2\delta}$, we have:

$$\begin{aligned} H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C}) = x^N) &= \sum_{i\in\mathcal{I}} H(Z_i|X_i = x_i) \\ &= \sum_{x\in\mathcal{X}} N(x|x_{\mathcal{I}})H(Z^*|X^* = x) \\ &\geq \sum_{x\in\mathcal{X}} N\alpha(1-2\delta)P_{X^*}(x)H(Z^*|X^* = x) \\ &= N\alpha(1-2\delta)H(Z^*|X^*), \end{aligned}$$

  where the function $N(x|x_{\mathcal{I}})$ represents the number of the letter x appearing in the sequence $x_{\mathcal{I}}$, and the inequality follows from the definition of $\tilde{T}_{\mathcal{I}}^N[X^*]_{2\delta}$. Therefore,

$$\begin{aligned} H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C})) &\geq H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C}), U_{\mathcal{I}}) \\ &\geq H(Z_{\mathcal{I}}^N(\mathcal{C})|X^N(\mathcal{C}), U_{\mathcal{I}} = 0)\Pr\{U_{\mathcal{I}} = 0\} \\ &\geq N\alpha(1-2\delta)(1-\epsilon_5)H(Z^*|X^*). \end{aligned} \tag{27}$$

By now, the terms in the rightmost side of Equation (21) have been bounded as expected. Substituting Equations (24) to (27) into (21) gives Equation (19). The proof of Lemma 3 is completed. □

4.4. Proofs of Equations (6) to (8)

Remark 7 and Lemma 3 yield:

$$\max_{\mathcal{I}\subseteq[1:N],|\mathcal{I}|=\mu} \frac{1}{N} I(W; Z_{\mathcal{I}}^N) < \epsilon$$

if the codebook is "good", which implies Equation (7). By the standard channel coding scheme (cf., for example, chapter 7.5 [28]), using the random codebook-generating scheme in Section 4.2, one can obtain a codebook satisfying Equation (8) with probability → 1 when $N \to \infty$. Combining Lemma 1, it is established that one can get a codebook achieving both equations of (7) and (8) with probability → 1. Equation (6) is obvious from the coding scheme. The proofs are completed.

5. Example

This section studies a concrete example of the communication model depicted in Figure 2 and formulated in Section 2, where the main channel is a discrete memoryless binary symmetrical channel (DM-MSC) with the crossover probability $0 \leq p \leq \frac{1}{2}$, and the eavesdropper observes arbitrary $\mu = N\alpha$

digital bits from the transmitter through a binary noiseless channel. This indicates that the transition matrices Q_M and Q_W (introduced in Definitions 2 and 3) satisfy that:

$$Q_M(y|x) = \begin{cases} p & \text{if } x \neq y, \\ 1-p & \text{otherwise} \end{cases}$$

and:

$$Q_W(z|x) = \begin{cases} 0 & \text{if } x \neq z, \\ 1 & \text{otherwise,} \end{cases}$$

where x, y and z all take the value from the binary alphabet $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$.

This example is in fact a special case of the model considered in [22]. Corollary 3 gives that the secrecy capacity of this example is:

$$C_s = C_s(\alpha, p) = \max_{U:U\to X\to Y} [I(U;Y) - \alpha I(U;X)], \tag{28}$$

where the auxiliary random variable U is sufficient to be binary. However, the formula given in (28) is inexplicit. In fact, it is an optimization problem. In this section, we will solve this optimization problem and find the exact random variable U achieving the "max" function. The main result is given in Proposition 1, whose proof is based on Lemma 4.

Suppose that the random variables of U, X and Y are all distributed on $\{0, 1\}$, and they satisfy that:

$$\begin{gathered} \Pr\{U=0\} = \beta_U, \Pr\{U=1\} = 1-\beta_U, \\ \Pr\{X=0|U=0\} = q_0, \Pr\{X=1|U=0\} = 1-q_0, \\ \Pr\{X=0|U=1\} = q_1, \Pr\{X=1|U=1\} = 1-q_1, \\ \Pr\{X=0\} = \beta_X = \beta_U \cdot q_0 + (1-\beta_U) \cdot q_1 \\ \text{and } \Pr\{X=1\} = 1-\beta_X, \end{gathered}$$

for some $0 \leq \beta_U, q_0, q_1 \leq 1$. Formula (28) can then be rewritten as:

$$C_s(\alpha, p) = \max_{0\leq\beta_U,q_0,q_1\leq 1} [h(\beta_X * p) - \alpha h(\beta_X) - \beta_U(h(q_0 * p) - \alpha h(q_0)) - (1-\beta_U)(h(q_1 * p) - \alpha h(q_1))],$$

where:

$$h(a) = -a\log a - (1-a)\log(1-a)$$

and:

$$a * b = a + b - 2ab$$

for $0 \leq a, b \leq 1$. Denoting:

$$g(\beta) = h(\beta * p) - \alpha h(\beta)$$

and:

$$C(\beta_U, q_0, q_1) = g(\beta_U q_0 + (1-\beta_U)q_1) - \beta_U g(q_0) - (1-\beta_U)g(q_1)$$

for $1 \leq \beta \leq 1$, the function $C_s(\alpha, p)$ can then be further represented as:

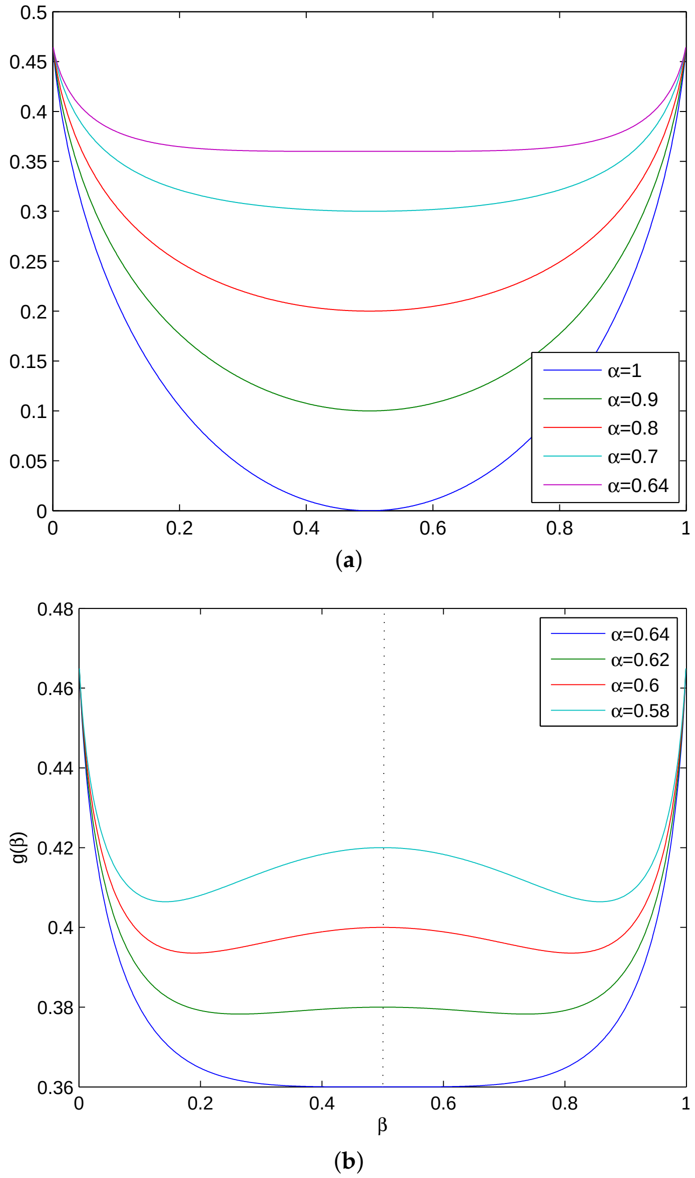
$$\begin{aligned} C_s(\alpha, p) &= \max_{0\leq\beta_U,q_0,q_1\leq 1} [g(\beta_X) - \beta_U g(q_0) - (1-\beta_U)g(q_1)] \\ &= \max_{0\leq\beta_U,q_0,q_1\leq 1} [g(\beta_U q_0 + (1-\beta_U)q_1) - \beta_U g(q_0) - (1-\beta_U)g(q_1)] \\ &= \max_{0\leq\beta_U,q_0,q_1\leq 1} C(\beta_U, q_0, q_1). \end{aligned}$$

To determine the value of $C_s(\alpha, p)$, some properties of the function g are given in the following lemma.

Lemma 4. *For any given $1 \leq \alpha, p \leq 1$, it follows that:*

1. *$g(\beta)$ is symmetrical around $\beta = \frac{1}{2}$.*
2. *when $\alpha \geq (1-2p)^2$, the function g is convex over $[0,1]$, and $\beta = \frac{1}{2}$ is the unique minimal point.*
3. *when $0 \leq \alpha < (1-2p)^2$, there exists a unique minimal point $\beta^* < \frac{1}{2}$ on the interval $[0, \frac{1}{2}]$, and hence, $1-\beta^*$ is the unique minimal point over the interval $[\frac{1}{2}, 1]$; moreover, $\beta = \frac{1}{2}$ is the unique maximal point over the interval $[\beta^*, 1-\beta^*]$, and the function is convex over the intervals of $[0, \beta^*]$ and $[1-\beta^*, 1]$.*

The proof of Lemma 4 is given in Appendix D. Figure 4 shows some examples on the function g, which cover both cases of $\alpha \geq (1-2p)^2$ and $\alpha < (1-2p)^2$.



(**a**)

(**b**)

Figure 4. Relationship between the function g and β with $p = 0.1$. (**a**) $\alpha \geq (1-2p)^2$; (**b**) $\alpha < (1-2p)^2$.

On account of Lemma 4, we conclude that:

Proposition 1. *The secrecy capacity can be represented as:*

$$C_S = 1 - \alpha - g(\beta^*), \tag{29}$$

where β^* is the unique minimal point of the function g over the interval $[0, \frac{1}{2}]$. Moreover, C_s is positive if and only if $\alpha < (1-2p)^2$.

Proof. When $\alpha \geq (1-2p)^2$, the function g is convex over the interval $[0, 1]$. Therefore:

$$C(\beta_U, q_0, q_1) \leq 0$$

for all β_U, q_0 and q_1. This indicates that $C_s = 0$. It remains to determine the capacity for the case of $\alpha < (1-2p)^2$. The inference is divided into two parts. The first part shows that the inequalities $\beta^* \leq \beta_X \leq 1 - \beta^*$ hold for all β_U, q_0 and q_1 satisfying $C(\beta_U, q_0, q_1) > 0$. The second part establishes Equation (29).

The first part is proven by contradiction. Suppose that $\beta_X \notin [\beta^*, 1-\beta^*]$. We can assume that $0 < \beta_X < \beta^*$ without loss of generality. In this case, it must follows that $q_0 \leq \beta_X \leq \beta^*$ or $q_1 \leq \beta_X \leq \beta^*$ since $\beta_X = \beta_U q_0 + (1-\beta_U) q_1$ is the convex combination of q_0 and q_1. We further suppose that $q_0 < \beta_X < \beta^*$. If it is also true that $q_1 < \beta^*$, then it follows immediately that $C(\beta_U, q_0, q_1) \leq 0$ on account of the fact that the function g is convex over the interval $[0, \beta^*]$. On the other hand, if $q_1 > \beta^*$, we let β_U^* satisfy that:

$$\beta_U^* q_0 + (1-\beta_U^*)\beta^* = \beta_X.$$

Then, it follows clearly that $\beta_U^* \leq \beta_U$, and hence:

$$\begin{aligned} C(\beta_U, q_0, q_1) &= g(\beta_U q_0 + (1-\beta_U) q_1) - \beta_U g(q_0) - (1-\beta_U) g(q_1) \\ &= g(\beta_U^* q_0 + (1-\beta_U^*) q^*) - \beta_U g(q_0) - (1-\beta_U) g(q_1) \\ &\overset{(a)}{\leq} g(\beta_U^* q_0 + (1-\beta_U^*) q^*) - \beta_U^* g(q_0) - (1-\beta_U^*) g(\beta^*) \\ &\overset{(b)}{\leq} 0, \end{aligned}$$

where (a) follows because β^* is the minimal point of g and $\beta_U^* < \beta_U$ and (b) follows because g is convex over the interval $[0, \beta^*]$. This contradicts the assumption that $C(\beta_U, q_0, q_1) > 0$.

To prove the second part, consider that when $\beta^* < \beta_X < 1 - \beta^*$, we have:

$$C(\beta_U, q_0, q_1) = g(\beta_X) - \beta_U g(q_0) - (1-\beta_U) g(q_1) \leq g(\frac{1}{2}) - g(\beta^*),$$

where the inequality holds because $\frac{1}{2}$ is the maximal point over the interval $[\beta^*, 1-\beta^*]$ and β^* is the minimal point over the interval $[0, 1]$. The formula above indicates that $C_s \leq g(\frac{1}{2}) - g(\beta^*) = 1 - \alpha - \beta^*$. The equality holds if $q_0 = \beta^*$, $q_1 = 1 - \beta^*$ and $\beta_U = \beta_X = \frac{1}{2}$; see Figure 5.

The proof of Proposition 1 is completed. □

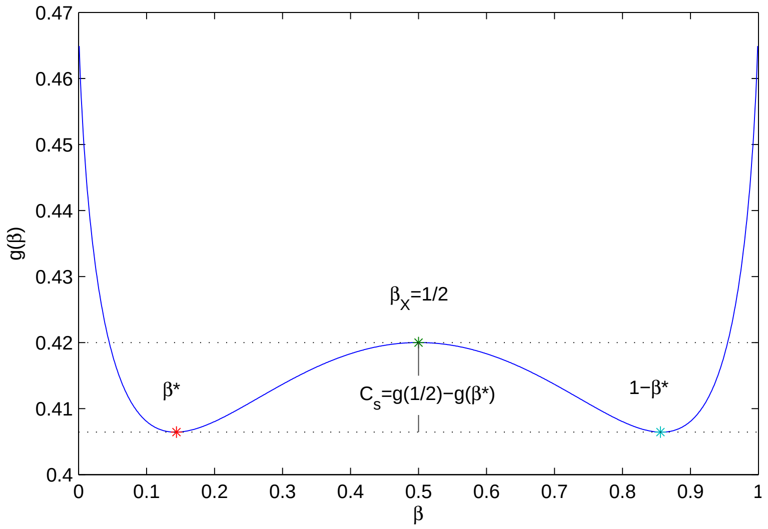


Figure 5. Parameters achieving the secrecy capacity ($\alpha = 0.58, p = 0.1$).

It is clear that when $p = 0$, the communication model discussed in this section is specialized as wiretap channel II. In that case, the secrecy capacity is obviously a linear function of α. However, when $p > 0$, the linearity does not hold. Instead, the secrecy capacity is a convex function of α. See Figure 6.

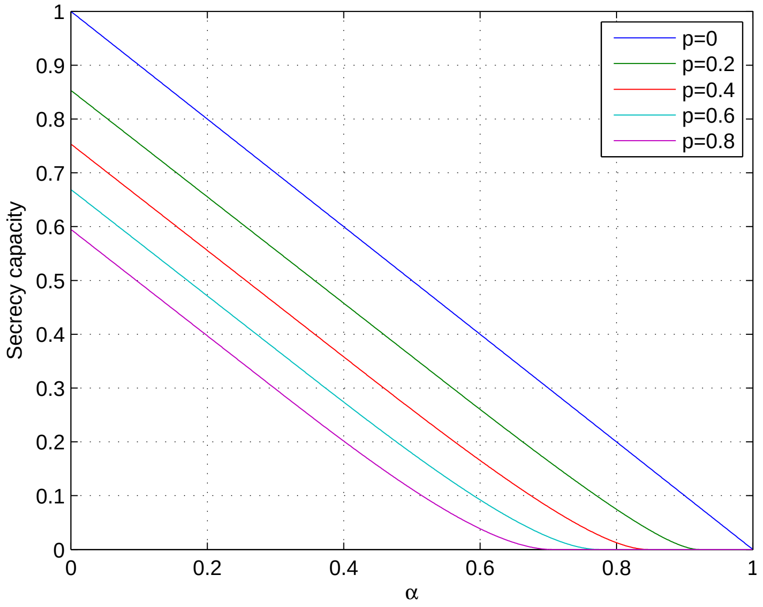


Figure 6. Relationship between the secrecy capacity $C_s(\alpha, p)$ and α.

6. Conclusions

The paper considers a communication model of extended wiretap channel II. In this new model, the source message is sent to the legitimate receiver through a discrete memoryless channel (DMC), and there exists an eavesdropper who is able to observe the digital sequence from the transmitter through a second DMC. The coding scheme is based on that developed by Ozarow and Wyner for the classic wiretap channel II. This communication model includes the general discrete memoryless wiretap channels, the wiretap channel II and the communication models discussed in [21–23] as special cases.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China under Grants 61271222, 61271174 and 61301178, the National Basic Research Program of China under Grant 2013CB338004 and the Innovation Program of Shanghai Municipal Education Commission under Grant 14ZZ017.

Author Contributions: Dan He and Wangmei Guo established the secrecy capacity of the binary example and wrote this paper. Dan He and Yuan Luo constructed the secure partition. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Coding Scheme Achieving Vanishing Maximal Decoding Error Probability

The coding scheme introduced in Section 4.2 can achieve the vanishing average decoding error probability, when the block length N is sufficiently large. This Appendix shows that the vanishing maximal decoding error probability could also be achieved by the standard techniques.

Appendix A.1. Preliminaries

The proof is based on the following two facts. The proof of Fact A can be found in Chapter 7.5 of [28], while Fact B can be obtained immediately.

Fact A: Let $\mathcal{C} = x^N(m)_{m=1}^{M'}$ be a deterministic codebook. Suppose that a deterministic coding scheme is applied to a given point-to-point noisy channel, i.e., the encoder is a bijection between the codebook and the message set. If the average decoding error probability of the codebook is ϵ, then there exists a sub-codebook $\tilde{\mathcal{C}}$ of the original codebook $\mathcal{C}$, such that $|\tilde{\mathcal{C}}| = M'/2$ and the maximal decoding error probability of the related deterministic coding scheme is no greater than 2ϵ.

Fact B: Let $\tilde{\mathcal{C}} = x^N(m)_{m=1}^{M'/2}$ be a deterministic codebook. Given a point-to-point noisy channel, suppose that the maximal decoding error probability of the related deterministic coding scheme is 2ϵ. Then, for any partition on that codebook, the random binning scheme introduced in Section 4.2 would achieve a maximal decoding error probability no greater than 2ϵ.

Appendix A.2. Coding Scheme

With the help of these two facts, the following coding scheme is achieved.

Codebook generation: Let $\mathcal{C}$ be a "good" codebook, such that the average decoding error probability (of the related deterministic coding scheme) is less than ϵ. Section 4.4 has shown the existence of this kind of codebook. On account of Fact A, there exists a sub-code $\tilde{\mathcal{C}}$ such that the maximal decoding error probability (of the related deterministic coding scheme) is less than 2ϵ. We will use $\tilde{\mathcal{C}}$ as the final codebook.

Codebook partition: Set $R = I(X^*;Y^*) - \alpha I(X^*;Z^*) - \tau$, $M = 2^{NR}$, and find a secure partition $\{\mathcal{C}_m\}_{m=1}^{M}$ on the codebook $\tilde{\mathcal{C}}$, such that:

$$\max_{\mathcal{I}\in\mathfrak{I}_\mu} \frac{1}{N} I(\tilde{W}; Z_{\mathcal{I}}^N(\tilde{\mathcal{C}})) < \epsilon,$$

where $\tilde{W}$ and $Z_{\mathcal{I}}^N(\tilde{\mathcal{C}})$ are similar to the corresponding random variables introduced in Section 4.2.

Encoder and decoder: This is similar to that introduced in Section 4.2.

Analysis of the maximal decoding error probability: It follows from Fact B that the a maximal decoding error probability of this coding scheme is no greater than 2ϵ.

Proof on the existence of the secure partition: By the property of the "good" codebook (see Definition 9), the codebook $\tilde{\mathcal{C}}$ satisfies that:

$$|\tilde{T}(\tilde{\mathcal{C}},\mathcal{I})| > (1-4\epsilon_1)\cdot\frac{M'}{2} \text{ for all } \mathcal{I}\in\mathfrak{I}_\mu$$

and:

$$|\tilde{T}(\tilde{\mathcal{C}},z^N,\mathcal{I})| < 2^{N(I(X^*;Y^*)-\alpha I(X^*;Z^*)-\tau-\frac{\tau_d}{2})} \text{ for all } \mathcal{I}\in\mathfrak{I}_\mu \text{ and } z^N \in \tilde{T}_{\mathcal{I}}^N[Z^*]_{2\delta}.$$

On account of Lemma 2, there exists a secure equipartition $\{\mathcal{C}_m\}_{m=1}^{M}$ on it, such that:

$$|\tilde{T}(\mathcal{C},\mathcal{I},z^N)\cap\mathcal{C}_m| < L$$

for all $1 \leq m \leq M, \mathcal{I} \in \mathfrak{I}_\mu$ and $z^N \in T^N_{\mathcal{I}}[Y^*]_{2\delta}$, where L is a sufficiently large constant. This partition is actually secure.

Appendix B. Proof of Lemma 1

According to the definition of a "good" codebook, it follows that:

$$\begin{aligned} \Pr\{\mathbf{C} \text{ is not "good" }\} \quad \leq \quad & \Pr\{\max_{\mathcal{I}\in\mathfrak{I}_\mu} \tilde{T}(\mathbf{C},\mathcal{I}) > 2\epsilon_1 M'\} \\ & + \Pr\{\max_{\mathcal{I}\in\mathfrak{I}_\mu} \max_{z^N\in\tilde{T}^N_{\mathcal{I}}[Z^*]_{2\delta}} \tilde{T}(\mathbf{C},z^N,\mathcal{I}) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\}. \end{aligned} \tag{B1}$$

On account of Lemma 4 in [23], the first term on the right-hand side of Equation (B1) is bounded by:

$$\Pr\{\min_{\mathcal{I}\in\mathfrak{I}_\mu} \tilde{T}(\mathbf{C},\mathcal{I}) < 2\epsilon_1 M'\} < \epsilon_3 = \exp_2\{N - \epsilon_1 M'\}.$$

Therefore, it suffices to prove that the second term on the right-hand side of Equation (B1) satisfies:

$$\Pr\{\max_{\mathcal{I}\in\mathfrak{I}_\mu} \max_{z^N\in\tilde{T}^N_{\mathcal{I}}[Z^*]_{2\delta}} \tilde{T}(\mathbf{C},z^N,\mathcal{I}) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\} < \epsilon_4.$$

To this end, denote:

$$U(i,z^N) = \begin{cases} 1 & \text{if } {X^*}^N_{\mathcal{I}}(i) \in T^N_{\mathcal{I}}[X^*Z^*|z^N]_{2\delta}, \\ 0 & \textit{otherwise} \end{cases}$$

for every $1 \leq i \leq M'$ and $z^N = \bigcup_{\mathcal{I}\in\mathfrak{I}_\mu} T^N_{\mathcal{I}}[Y^*]_{2\delta}$. Then, on account of the Chernoff bound, it follows that:

$$\begin{aligned} & \Pr\{\tilde{T}(\mathbf{C},z^N,\mathcal{I}) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\} \\ = \quad & \Pr\{\sum_{i=1}^{M'} U(i,z^N) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\} \\ < \quad & \exp_2(-2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}) \cdot E[\exp_2(\sum_{i=1}^{M'} U(i,z^N))] \\ < \quad & \exp_2(-2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}) \cdot \prod_{i=1}^{M'} E[\exp_2(U(i,z^N))] \\ < \quad & \exp_2(-2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}) \cdot \prod_{i=1}^{M'} \exp_e(E[U(i,z^N)]), \end{aligned} \tag{B2}$$

where the last inequality follows because $2^t \leq 1 + t \leq e^t$ for $0 \leq t \leq 1$. Recalling Corollary 5, we have:

$$E[U(i,z^N) < 2^{-N(\alpha I(X^*;Z^*)-4\delta H(X^*))}.$$

Substituting the formula above to (B2) gives:

$$\begin{aligned} & \Pr\{\tilde{T}(\mathbf{C},z^N,\mathcal{I}) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\} \\ < \quad & \exp_2(-2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}) \cdot \prod_{i=1}^{M'} \exp_e(E[U(i,z^N)]) \\ < \quad & \exp_2(-2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}) \cdot \exp_e(2^{-N(\alpha I(X^*;Z^*)-4\delta H(X^*))} M') \\ = \quad & \exp_2[(2^{-N(\tau_d-4\delta H(X^*))} \log e - 2^{-\frac{1}{2}N\tau_d}) 2^{N(I(X^*;Y^*)-\alpha I(X^*;Z^*)-\tau)}] \\ = \quad & \exp_2[(2^{-N(\tau_d-4\delta H(X^*))} \log e - 2^{-\frac{1}{2}N\tau_d}) M], \end{aligned}$$

where M' and M are given by Equations (10) and (11), respectively. Therefore,

$$\begin{aligned}
& \Pr\{\max_{\mathcal{I}\in\mathfrak{I}_\mu} \max_{z^N\in\tilde{T}^N_{\mathcal{I}}[Z^*]_{2\delta}} \tilde{T}(\mathbf{C},z^N,\mathcal{I}) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\} \\
\leq & \sum_{\mathcal{I}\in\mathfrak{I}_\mu} \sum_{z^N\in\tilde{T}^N_{\mathcal{I}}[Z^*]_{2\delta}} \Pr\{\tilde{T}(\mathbf{C},z^N,\mathcal{I}) > 2^{N(I(X^*;Y^*)-\alpha I(X^*;Y^*)-\tau-\frac{\tau_d}{2})}\} \\
\leq & \sum_{\mathcal{I}\in\mathfrak{I}_\mu} \sum_{z^N\in\tilde{T}^N_{\mathcal{I}}[Z^*]_{2\delta}} \exp_2[(2^{-N(\tau_d-4\delta H(X^*))}\log e - 2^{-\frac{1}{2}N\tau_d})M] \\
\leq & \exp_2[2N + (2^{-N(\tau_d-4\delta H(X^*))}\log e - 2^{-\frac{1}{2}N\tau_d})M] \\
= & \epsilon_4.
\end{aligned}$$

The proof is completed.

Appendix C. Proof of Lemma 2

This Appendix proves that for any given “good” codebook $\mathcal{C}$, there exists a secure partition $\{\mathcal{C}_m\}_{m=1}^M$ on it such that:

$$|\tilde{T}(\mathcal{C},\mathcal{I},z^N)\cap\mathcal{C}_m| < L$$

for all $1 \leq m \leq M$, $\mathcal{I} \in \mathfrak{I}_\mu$ and $z^N \in T^N_{\mathcal{I}}[Y^*]_{2\delta}$. The proof is quite similar to the proof of Lemma 2 in [14]. Most notation in this Appendix will follow that in [14].

Let $\mathcal{F}$ be the set of all possible equipartition on the codebook $\mathcal{C}$. Each element in $\mathcal{F}$ is actually a function $f : \mathcal{C} \mapsto [1 : M]$ such that:

$$|f^{-1}(m)| = r = \frac{M'}{M} = 2^{N(\alpha I(X^*;Z^*)-\tau_d)}.$$

For any $f \in \mathcal{F}$, let $\Psi(f) = 0$ if the partition produced by f is secure, and $\Psi(f) = 1$ otherwise. Then, it suffices to prove that:

$$E[\Psi(F)] < 1,$$

where F is the random variable uniformly distributed on $\mathcal{F}$. To this end, for any $1 \leq m \leq M$, denote:

$$\Phi(f,m,\mathcal{I},z^N) = \begin{cases} 0 & \text{if } |\tilde{T}(\mathcal{C},\mathcal{I},z^N)\cap f^{-1}(m)| < L, \\ 1 & \text{otherwise.} \end{cases}$$

It follows clearly that:

$$E[\Psi(F)] \leq \sum_{m=1}^{M} \sum_{\mathcal{I}\in\mathfrak{I}_\mu} \sum_{z^N\in T^N_{\mathcal{I}}[Y^*]_{2\delta}} E[\Phi(F,m,\mathcal{I},z^N)]. \tag{C1}$$

In the remainder of the proof, we firstly bound the value of $E[\Phi(F,m,\mathcal{I},z^N)]$, and then, the value of $E[\Psi(F)]$ is bounded by Equation (C1).

Upper bound of $E[\Phi(F,m,\mathcal{I},z^N)]$. For any $\mathcal{I} \in \mathfrak{I}_\mu$ and $z^N \in T^N_{\mathcal{I}}[Y^*]_{2\delta}$, let:

$$n_{z^N} = |\tilde{T}(\mathcal{C},\mathcal{I},z^N)|.$$

Then, it follows that:

$$n_{z^N} \leq n_1 = 2^{N(I(X^*;Y^*)-\alpha I(X^*;Z^*)-\tau-\frac{\tau_d}{2})}.$$

since the codebook $\mathcal{C}$ is "good". Therefore, the probability that there are t codewords of $F^{-1}(m)$ belonging to $\tilde{T}(\mathcal{C}, \mathcal{I}, z^N)$ is given by:

$$\Pr\{|\tilde{T}(\mathcal{C}, \mathcal{I}, z^N) \cap F^{-1}(m)| = t\} = \frac{\binom{n_{z^N}}{t}\binom{M'-n_{z^N}}{r-t}}{\binom{M'}{r}} \le \frac{\binom{n_1}{t}\binom{M'}{r-t}}{\binom{M'}{r}}.$$

By the similar method of bound π_t in [14], we have:

$$\binom{n_1}{t} \cdot \frac{\binom{M'}{r-t}}{\binom{M'}{r}} \le \frac{n_1^t}{t!} \cdot \frac{r!(M'-r)!}{(r-t)!(M'-r+1)!} = \frac{n_1^t}{t!} \cdot \sum_{j=1}^{t} \frac{r-j+1}{M'-r+j} \le \frac{n_1^t}{t!} \cdot \frac{r^t}{(M'-r)^t} = \frac{(n_1 r/M')^t}{t!(1-r/M')^t}.$$

Observing that $\frac{n_1 r}{M'} = 2^{-N\tau_d/2}$ and $1 - \frac{r}{M'} > \frac{1}{2}$, we have:

$$\Pr\{|\tilde{T}(\mathcal{C}, \mathcal{I}, z^N) \cap F^{-1}(m)| = t\} \le 2^{-N\tau_d t/2}\frac{2^t}{t!}.$$

This indicates that:

$$\begin{aligned} E[\Phi(F, m, \mathcal{I}, z^N)] &= \sum_{t=L}^{r} \Pr\{|\tilde{T}(\mathcal{C}, \mathcal{I}, z^N) \cap F^{-1}(m)| = t\} \\ &\le \sum_{t=L}^{r} 2^{-N\tau_d t/2}\frac{2^t}{t!} \\ &\le \sum_{t=L}^{\infty} 2^{-N\tau_d L/2}\frac{2^t}{t!} \\ &= e^2 \cdot 2^{-N\tau_d L/2}. \end{aligned} \tag{C2}$$

Upper bound of $E[\Psi(F]$. Substituting Equation (C2) into (C1) gives:

$$E[\Psi(F)] \le \sum_{m=1}^{M} \sum_{\mathcal{I} \in \mathfrak{I}_\mu} \sum_{z^N \in T_{\mathcal{I}}^N[Y^*]_{2\delta}} e^2 \cdot 2^{-N\tau_d L/2}.$$

Observing that $M = 2^{NR}$, $|\mathfrak{I}_\mu| \le 2^N$ and $T_{\mathcal{I}}^N[Y^*]_{2\delta} \le 2^{N\alpha}$ for $\mathcal{I} \in \mathfrak{I}_\mu$, the formula above can be further bounded by:

$$E[\Psi(F)] \le e^2 \cdot 2^{N(R+1+\alpha-\tau_d L/2)},$$

which is < 1 if $L \ge 2(R+6)/\tau_d$. This completes the proof of Lemma 2.

Appendix D. Proof of Lemma 4

This Appendix establishes the properties of the function g given in Lemma 4. Property 1 is obvious. We only give the proof for Properties 2 and 3.

Proof of Property 2. The first and the second derivatives of the function g are:

$$g'(\beta) = \log e[(1-2p)\ln\frac{1-\beta * p}{\beta * p} - \alpha \ln\frac{1-\beta}{\beta}]$$

and:

$$g''(\beta) = \log e[\alpha(\frac{1}{\beta} + \frac{1}{1-\beta}) - (1-2p)^2(\frac{1}{\beta * p} + \frac{1}{1-\beta * p})],$$

respectively. For any $0 \le \beta \le \frac{1}{2}$, it follows that $\beta \le \beta * p \le \frac{1}{2}$. Therefore, when $\alpha \ge (1-2p)^2$, we have:

$$g''(\beta) > 0$$

for $0 \leq \beta \leq \frac{1}{2}$. Moreover, noticing that:

$$g'(\frac{1}{2}) = 0,$$

we conclude that:

$$g'(\beta) \leq 0$$

for $0 \leq \beta \leq \frac{1}{2}$. This implies that the function g is decreasing and convex over the interval $[0, \frac{1}{2}]$. On account of the symmetry, it is concluded that g is increasing and convex over the interval $[\frac{1}{2}, 1]$, and hence, $\beta = \frac{1}{2}$ is the unique minimal point in the interval $[0, 1]$. One can easily verify that $\beta = \frac{1}{2}$ is actually a convex point. Thus, the function is convex over the whole interval $[0, 1]$. □

Proof of Property 3. We firstly find the solution for the inequality $g''(\beta) < 0$. This inequality indicates that:

$$\frac{\frac{1}{\beta} + \frac{1}{1-\beta}}{\frac{1}{\beta * p} + \frac{1}{1-\beta * p}} = \frac{(\beta * p)(1 - \beta * p)}{\beta(1-\beta)} < \frac{(1-2p)^2}{\alpha},$$

or:

$$\beta^2 - \beta + \frac{p - p^2}{(1-2p^2)(1/\alpha - 1)} < 0.$$

The inequality further indicates that $\beta'' < \beta < 1 - \beta''$ for some $0 < \beta'' < \frac{1}{2}$.

Then, we study the property of g' over the interval $[0, \frac{1}{2}]$. It is clear that g' increases on the interval $[0, \beta'']$ and decreases on the interval $[\beta'', \frac{1}{2}]$. Furthermore, since $g'(\frac{1}{2}) = 0$, it follows that $g(\beta'') > 0$. Noticing the facts that g' increases over the interval $[0, \beta'']$ and $g'(0) < 0$, we conclude that there exists a unique point $0 < \beta^* < \beta''$ such that $g'(\beta^*) = 0$, and that point is clearly a unique minimal point over the interval $[0, \frac{1}{2}]$.

Finally, since $\beta^* < \beta''$, it follows that $g''(\beta) > 0$ for $\beta \in [0, \beta^*]$, and the function is convex over that interval. The property of g over the interval $[\frac{1}{2}, 1]$ is from the symmetry. It is also clear that $\beta = \frac{1}{2}$ is the unique maximal point over the interval $[\beta^*, 1 - \beta^*]$.

This completes the proof of this lemma. □

References

1. Wyner, A.D. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387.
2. Csiszár, I.; Körner, K. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348.
3. Chen, Y.; Vinck Han, A.J. Wiretap channel with side information. *IEEE Trans. Inf. Theory* **2008**, *54*, 395–402.
4. Dai, B.; Luo, Y. Some new results on the wiretap channel with side information. *Entropy* **2012**, *14*, 1671–1702.
5. Dai, B.; Vinck Han, A.J.; Luo, Y.; Zhuang, Z. Degraded Broadcast Vhannel with Noncausal Side Information, Confidential Messages and Noiseless Feedback. In Proceedings of the IEEE International Symposium on Information Theory, Cambridge, MA, USA, 1–6 July 2012.
6. Khisti, A.; Diggavi, S.N.; Womell, G.W. Secrete-key agreement with channel state information at the transmitter. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 672–681.
7. Chia, Y.H.; El Gamal, A. Wiretap channel with causal state information. *IEEE Trans. Inf. Theory* **2012**, *58*, 2838–2849.
8. Dai, B.; Luo, Y.; Vinck Han, A.J. Capacity Region of Broadcast Channels with Private Message and Causual Side Information. In Proceedings of the 3rd International Congress on Image and Signal Processing, Yantai, China, 16–18 October 2010.
9. Liang, Y.; Kramer, G.; Poor, H.; Shamai, S. Compound wiretap channel. *EURASIP J. Wirel. Commun. Netw.* **2008**, doi:10.1155/2009/142374.
10. Bjelaković, I.; Boche, H.; Sommerfeld, J. Capacity results for compound wiretap channels. *Probl. Inf. Transm.* **2011**, *49*, 73–98.
11. Schaefer, R.F.; Boche, H. Robust broadcasting of common and confidential messages over compound channels: Strong secrecy and decoding performance. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1720–1732.

12. Schaefer, R.F.; Loyka, S. The secrecy capacity of compound gaussian MIMO wiretap channels. *IEEE Trans. Inf. Theory* **2015**, *61*, 5535–5552.
13. Bjelaković, I.; Boche, H.; Sommerfeld, J. Capacity results for arbitrarily varying wiretap channel. *Inf. Theory Comb. Search Theory Lect. Notes Comput. Sci.* **2013**, *7777*, 123–144.
14. Ozarow, L.H.; Wyner, A.D. Wire-tap channel II. *AT&T Bell Labs Tech. J.* **1984**, *63*, 2135–2157.
15. Cai, N.; Yeung, R.W. Secure Network Coding. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Lausanne, Switzerland, 30 June–5 July 2002.
16. Cai, N.; Yeung, R.W. Secure network coding on a wiretap network. *IEEE Trans. Inf. Theory* **2011**, *57*, 424–435.
17. Feldman, J.; Malkin, T.; Stein, C. On the Capacity of Secure Network Coding. In Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2004.
18. Bhattad, K.; Narayanan, K.R. Weakly secure network coding. *Proc. NetCod* **2005**, *4*, 1–5.
19. Cheng, F.; Yeung, R.W. Performance bounds on a wiretap network with arbitrary wiretap sets. *IEEE Trans. Inf. Theory* **2012**, *60*, 3345–3358.
20. He, D.; Guo, W. Strong secrecy capacity of a class of wiretap networks. *Entropy* **2016**, *18*, 238.
21. Nafea, M.; Yener, A. Wiretap Channel II with a Noisy Main Channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015.
22. Goldfeld, Z.; Cuff, P.; Permuter, H.H. Semantic-security capacity for wiretap channels of type II. *IEEE Trans. Inf. Theory* **2016**, *62*, 3863–3879.
23. He, D.; Luo, Y.; Cai, N. Strong Secrecy Capacity of the Wiretap Channel II with DMC Main Channel. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016.
24. Luo, Y.; Mitrpant, C.; Vinck Han, A.J. Some new characters on the wiretap channel of type II. *IEEE Trans. Inf. Theory* **2005**, *51*, 1222–1229.
25. Yang, Y.; Dai, B. A combination of the wiretap channel and the wiretap channel of type II. *J. Comput. Inf. Syst.* **2014**, *10*, 4489–4502.
26. He, D.; Luo, Y. A Kind of Non-DMC Erasure Wiretap Channel. In Proceedings of the IEEE International Conference on Communication Technology (ICCT), Chengdu, China, 9–11 November 2012.
27. Kramer, G. Topics in multi-user information theory. Foundations Trends®. *Commun. Inf. Theory* **2007**, *4*, 265–444.
28. Yeung, R.W. *Information Theory and Network Coding*; Springer: New York, NY, USA, 2008.