Secure multiplex coding attaining channel capacity in wiretap channels

Daisuke Kobayashi, Hirosuke Yamamoto, *Fellow, IEEE,* and Tomohiro Ogawa

***Abstract*—It is known that a message can be transmitted safely against any wiretapper via a noisy channel without a secret key if the coding rate is less than the so-called secrecy capacity C_S, which is usually smaller than the channel capacity C. In order to remove the loss $C - C_S$, we propose a multiplex coding scheme with plural independent messages. In this paper, it is shown that the proposed multiplex coding scheme can attain the channel capacity as the total rate of the plural messages and the perfect secrecy for each massage. Several bounds of achievable multiplex coding rate region are derived for general wiretap channels in the sense of Information-Spectral methods, by extending Hayashi's proof, in which the coding of the channel resolvability is applied to wiretap channels. Furthermore, the exact region for deterministic coding is determined for stationary memoryless full-rank wiretap channels.**

***Index Terms*—wiretap channel, channel resolvability, information-spectrum method, secrecy capacity, multiplex coding**

I. Introduction

When Alice sends a message to Bob via a public noisy channel, Eve may wiretap the message. But, since the main channel from Alice to Bob has usually a different characteristic from the wiretap channel from Alice to Eve, we can devise a code such that the perfect secrecy against Eve can be attained without any secret key. The maximum attainable coding rate of such code is called the *secrecy capacity* C_S, which is generally smaller than the channel capacity C of the main channel.

The coding problem for the wiretap channels was first studied by Wyner [1]. Although the main and wiretap channels can be considered as a kind of broadcast channel [2], Wyner proved the coding theorem for the case of the so-called degraded broadcast channel. For more general broadcast channels, Csiszár and Körner [3] proved that the secrecy capacity C_S is given by $\max_{\widetilde{X}\to X\to(Y,Z)}[I(\widetilde{X};Y)-I(\widetilde{X};Z)]$, where X is the input, and Y and Z are the output of the main and wiretap channels, respectively. $\widetilde{X}$ is an auxiliary random variable that makes a Markov chain $\widetilde{X} \to X \to (Y, Z)$. In order to achieve the perfect secrecy with a positive coding rate, the channels must satisfy that $I(\widetilde{X};Y) > I(\widetilde{X};Z)$ for some $\widetilde{X}$, i.e., the main channel must be less noisy than the wiretap channel in this sense. But, even in the case that the main channel is more noisy than the wiretap channel, Maurer [4] devised a protocol which can attain the perfect secrecy if a public noiseless channel can be used.

In the above studies, channels are assumed to be stationary and memoryless. On the contrary, the information-spectrum methods [5] have been developed, and many kinds of coding theorems have been proved for the so-called general sources and general channels, which might be neither Ergodic nor stationary. As one of them, Han and Verdú [6] studied the so-called channel resolvability problem, in which we want to approximate the output probability distribution of a noisy channel for a given input probability distribution by encoding a random number. The minimum rate of the random number necessary to attain the approximation is called the channel resolvability, and they developed the theory of the channel resolvability for general channels. We note that one of their motivations to study the channel resolvability problem was to prove the converse part [6] of the theorem for identification coding [7], where the random variables arising as the output and the input of the channel may be neither Ergodic nor stationary.

On the other hand, for quantum channels, Devetak [8] introduced a stochastic encoder to realize a non-distinguishable probability distribution for any message at the output of a wiretap channel. Based on these background, Hayashi [9] considered the coding problem of general wiretap channels in the framework of the stochastic encoders and the channel resolvability, and established the method to prove the coding theorem of general wiretap channels. In addition to the generality, Hayashi's method has simplicity that we can divide wiretap channel coding into two viewpoints: ordinary message transmission coding and channel resolvability coding, although the description by the information-spectrum methods is almost inevitable from the history of the channel resolvability problem.

It is well known from the coding theorems proved in the previous studies that messages cannot be transmitted at any rate larger than the secrecy capacity C_S if we want to attain the perfect secrecy. Since C_S is generally less than the channel capacity C, we must loss $C - C_S$ in exchange for the secrecy. But, in this paper, we will devise multiplex coding of plural independent messages to remove the loss, and we will show that the channel capacity C can be attained as the total rate of the plural messages and each message can be protected with the perfect secrecy.

To prove the coding theorems for the multiplex coding scheme, we utilize Hayashi's method in wiretap channel coding, which has an advantage to enjoy simplicity described above. For general wiretap channels in the sense of

D. Kobayashi is with NTT Data Corporation, Koto-ku, Tokyo, 135-8671 Japan (email: kobayashidib@nttdata.co.jp)

H. Yamamoto is with the School of Frontier Sciences, The University of Tokyo Kashiwa-shi, Chiba, 277-8561 Japan (email: Hirosuke@ieee.org)

T. Ogawa is with the School of Information Systems, The University of Electro-Communications, Chofu-shi, Tokyo, 182-8585 Japan (email: ogawa@is.uec.ac.jp)

Information-Spectral methods, several bounds of achievable multiplex coding rate region are derived for the cases that encoding is deterministic or stochastic, and for the cases that security is measured by normalized mutual information $I(K_t;\mathbf{Z})/n$ between each secret message K_t and wiretap channel output $\mathbf{Z}$, where n is the code length, or by the average or maximum variational distance among wiretap channel output distributions for secret messages. Furthermore, for stationary memoryless full-rank wiretap channels, the exact achievable rate region of multiplex deterministic encoding is determined for the cases that security is measured by mutual information $I(K_t;\mathbf{Z})$ and the average or maximum variational distances.

In Section II, we define several technical terms, which are used in the information-spectrum methods. The multiplex coding scheme of plural messages is proposed in Section III. The main theorems are also shown in Section III although they are proved in Section IV. The case of stationary memoryless wiretap channels is treated in Section V. Finally it is shown in Section VI how the multiplex coding can be realized by linear coding for binary symmetric wiretap channels in the case of security measure $I(K_t;\mathbf{Z})/n$. In this paper, both input and output alphabets of channels are assumed to be discrete.

II. Preliminaries

According to the information-spectrum methods [5], a general random process, which might be neither Ergodic nor stationary, is denoted by

$$\mathbf{X} = \{X^n\}_{n=1}^{\infty}, \tag{1}$$

where $X^n = (X_1, X_2, \cdots, X_n)$ and each X_i, $i = 1, 2, \cdots, n$, takes values in discrete alphabets $\mathcal{X}^n$ and $\mathcal{X}$, respectively, and $\mathcal{X}^n$ is the Cartesian product of $\mathcal{X}$.

For two general random process $\mathbf{X}$ and $\mathbf{Y}$, the spectral sup-mutual information rate and the spectral inf-mutual information rate are defined as follows.

Definition 1: Spectral sup-mutual information rate:

$$\overline{I}(P_{\mathbf{X}}, P_{\mathbf{Y}|\mathbf{X}}) \equiv \inf\left\{\alpha \;\middle|\; \lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} > a\right\} = 0\right\}. \tag{2}$$

Spectral inf-mutual information rate:

$$\underline{I}(P_{\mathbf{X}}, P_{\mathbf{Y}|\mathbf{X}}) \equiv \sup\left\{\beta \;\middle|\; \lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} < \beta\right\} = 0\right\}. \tag{3}$$

Remark 1: In the case that $\mathbf{X}$ and $\mathbf{Y}$ are i.i.d. processes with a probability distribution $P_{X,Y}$, both of the spectral sup- and inf-mutual information rates coincide with the ordinary mutual information $I(X;Y)$ from the law of large numbers. In the following, readers who are not familiar to the information-spectrum methods may regard the general processes $\mathbf{X}$, $\mathbf{Y}$ as the i.i.d. processes and the general channels defined below as stationary memoryless channels, and so on.

A general channel $\mathbf{W}$ with an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$ is defined as $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$, where $W^n(\cdot|\cdot)$ is an arbitrary conditional probability distribution that satisfies

$$\sum_{y^n\in\mathcal{Y}^n} W^n(y^n|x^n) = 1 \tag{4}$$

for each $x^n \in \mathcal{X}^n$ and each $n = 1, 2, \cdots$. For the input process $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ and the output process $\mathbf{Y} = \{Y^n\}_{n=1}^{\infty}$ of the general channel $\mathbf{W}$, W^n satisfies for any $n > 0$ that

$$P_{X^n,Y^n}(x^n, y^n) = P_{X^n}(x^n)W^n(y^n|x^n), \tag{5}$$

$$P_{Y^n|X^n}(y^n|x^n) = W^n(y^n|x^n), \tag{6}$$

$$P_{Y^n}(y^n) = \sum_{x^n\in\mathcal{X}^n} P_{X^n}(x^n)W^n(y^n|x^n) \equiv P_{X^n}W^n(y^n). \tag{7}$$

Note that P_{Y^n} is also denoted by $P_{X^n}W^n$ because P_{Y^n} is determined by P_{X^n} and W^n.

For simplicity, the alphabets of a general channel are denoted by $\mathcal{X} \to \mathcal{Y}$ when the input and output alphabets are $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\mathbf{U}$ and $\mathbf{W}$ be general channels with alphabets $\widetilde{\mathcal{X}} \to \mathcal{X}$ and $\mathcal{X} \to \mathcal{Y}$, respectively. Then, the cascade channel $\mathbf{UW}$ with alphabets $\widetilde{\mathcal{X}} \to \mathcal{Y}$ is defined by $\mathbf{UW} = \{(UW)^n\}_{n=1}^{\infty}$, where

$$(UW)^n(y^n|\widetilde{x}^n) \equiv \sum_{x^n\in\mathcal{X}^n} U^n(x^n|\widetilde{x}^n)W^n(y^n|x^n). \tag{8}$$

Now we consider the channel resolvability problem. Let $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ be a general channel with alphabets $\mathcal{X} \to \mathcal{Z}$, and let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ be an input and $\mathbf{Z} = \{Z^n\}_{n=1}^{\infty}$ be the corresponding output. Then, we want to approximate the output $\mathbf{Z}$ by inputting $\widehat{\mathbf{X}} = \{\widehat{X}^n\}_{n=1}^{\infty}$ into the channel, where $\widehat{X}^n$ is generated by encoding a uniform random number K over an alphabet $\mathcal{K} \equiv \{1, 2, \cdots, M_{(n)}\}$. For the output $\widehat{\mathbf{Z}} = \{\widehat{Z}^n\}_{n=1}^{\infty}$ of the input $\widehat{\mathbf{X}}$, we evaluate the performance of the approximation between $\mathbf{Z}$ and $\widehat{\mathbf{Z}}$ by the variational distance $d(Z^n, \widehat{Z}^n) = \left\|P_{Z^n} - P_{\widehat{Z}^n}\right\|_1 = \sum_{z^n} |P_{Z^n}(z^n) - P_{\widehat{Z}^n}(z^n)|$.

Definition 2: For a given channel $\mathbf{V} = \{V^n\}_{n=1}^{\infty}$ with alphabets $\mathcal{X} \to \mathcal{Y}$, a rate R is called achievable for an input $\mathbf{X}$ if there exists a sequence of codes $\varphi_n : \widehat{X}^n = \varphi_n(K)$ that satisfies

$$\lim_{n\to\infty} d(Z^n, \widehat{Z}^n) = 0, \tag{9}$$

$$\limsup_{n\to\infty} \frac{1}{n}\log M_{(n)} \le R, \tag{10}$$

where $P_{Z^n}(z^n) = P_{X^n}V^n(z^n)$, $P_{\widehat{Z}^n}(z^n) = P_{\widehat{X}^n}V^n(z^n)$, and K is the uniform random number over $\mathcal{K} \equiv \{1, 2, \cdots, M_{(n)}\}$. Furthermore, the channel resolvability for an input $\mathbf{X}$, denoted by $S_{\mathbf{X}}(\mathbf{V})$, is defined as follows.

$$S_{\mathbf{X}}(\mathbf{V}) \equiv \inf\{R \mid R \text{ is achievable for the input } \mathbf{X} \text{ of the channel } \mathbf{V}\} \tag{11}$$

Then, Han-Verdú [6] proved the next theorem.

Theorem 1: For any general channel $\mathbf{V}$ and any input $P_{\mathbf{X}}$, it holds that

$$S_{\mathbf{X}}(\mathbf{V}) \le \overline{I}(P_{\mathbf{X}}, \mathbf{V}). \tag{12}$$

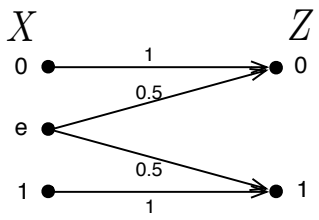

Fig. 1: A non-full-rank channel.

Remark 2: The equality of (12) does not always hold. For instance[1], consider a stationary memoryless channel $V(z|x)$ given by Fig. 1 with $P_X(0) = P_X(1) = 0.5$. In this case, we have $P_Z(0) = P_Z(1) = 0.5$ and $\overline{I}(P_{\mathbf{X}}, \mathbf{V}) = I(X;Z) = \log 2$ while this P_Z can be realized with $S_{\mathbf{X}}(\mathbf{V}) = 0$ by using $P_{\widehat{X}}(\mathrm{e}) = 1$. But, it is shown in [10] that if $\mathbf{V}$ is a full-rank channel, i.e., $\mathbf{V}$ is a stationary memoryless channel such that $\{V(\cdot|x)\}$, $x \in \mathcal{X}$, are linearly independent as a set of vectors, then the channel resolvability $S_{\mathbf{X}}(\mathbf{V})$ satisfies that for any input $\mathbf{X}$,

$$S_{\mathbf{X}}(\mathbf{V}) = \overline{I}(P_{\mathbf{X}}, \mathbf{V}). \tag{13}$$

Note that Fig. 1 is not a full rank channel because every full-rank channel satisfies $|\mathcal{X}| \le |\mathcal{Z}|$.

III. Multiplex Coding

In this section, we consider multiplex coding for wiretap channels. Assume that Alice sends messages to Bob via a main channel $\mathbf{W}$ with alphabets $\mathcal{X} \to \mathcal{Y}$ and Eve eavesdrops the messages via a wiretap channel $\mathbf{V}$ with alphabets $\mathcal{X} \to \mathcal{Z}$. $\mathbf{W}$ and $\mathbf{V}$ are general channels in the sense of the information-spectrum methods. The input of both channels is $P_{\mathbf{X}}$, and the outputs of $\mathbf{W}$ and $\mathbf{V}$ are $P_{\mathbf{Y}}$ and $P_{\mathbf{Z}}$, respectively.

Assume that Alice sends T independent messages $K_1, K_2, \cdots, K_T$ to Bob by multiplex coding. Each K_t, $t = 1, 2, \cdots, T$, takes values in $\mathcal{K}_t \equiv \{1, 2, \cdots, M_t\}$, and satisfies that $\Pr\{K_t = k\} = 1/M_t$ for all $k \in \mathcal{K}_t$. The aim of the multiplex coding is to attain the following performance.

(A) Every K_t must be transmitted to Bob within an arbitrarily small error probability.

(B) The perfect secrecy against Eve must be attained for each K_t, $t = 1, 2, \cdots, T$, individually.

Note that the above (B) does not require the perfect secrecy of the entire $(K_1, K_2, \cdots, K_T)$, which is usually required in the ordinary (i.e. non-multiplex) coding for wiretap channels. In the case of (B), although any information about each K_t does not leak out, some information about the combination of $(K_1, K_2, \cdots, K_T)$ may leak out. But, since K_t, $t = 1, 2, \cdots, T$, are assumed to be mutually independent, the combination has no meaning, and hence, the individual perfect secrecy of K_t is reasonable. (See Remark 5 for further discussions.)

The tuple $(K_1, K_2, \cdots, K_T)$ is encoded by an encoder φ_n to a codeword X^n, which is sent to Bob via the main channel W^n. In this paper, we consider the case that stochastic encoders can be used in addition to the case that only deterministic encoders can be used. Formally, a deterministic encoder is described by a map

$$\varphi_n : (k_1, k_2, \cdots, k_T) \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_T \longmapsto x^n_{k_1,k_2,\cdots,k_T} \in \mathcal{X}^n, \tag{14}$$

while a stochastic encoder is described by

$$\varphi_n : (k_1, k_2, \cdots, k_T) \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_T \longmapsto Q_{k_1,k_2,\cdots,k_T} \in \mathcal{P}(\mathcal{X}^n), \tag{15}$$

where $\mathcal{P}(\mathcal{X}^n)$ is the set of probability distributions on the set $\mathcal{X}^n$. In the case of the stochastic encoders, an input X^n is generated according to the probability distribution $Q_{k_1,k_2,\cdots,k_T}$ when the tuple of messages is $(k_1, k_2, \cdots, k_T)$. Then, the input X^n yields the output Y^n for Bob via the main channel W^n, while it yields the output Z^n for Eve via the wiretap channel V^n.

The description of the deterministic encoder (14) is unified into that of the stochastic encoder (15) using the point mass distribution:

$$Q_{k_1,k_2,\cdots,k_T}(x^n) = \delta_{\varphi_n(k_1,k_2,\cdots,k_T)}(x^n) \tag{16}$$

where

$$\delta_{\varphi_n(k_1,k_2,\cdots,k_T)}(x^n) \equiv \begin{cases} 1 & \text{if} \quad x^n = \varphi_n(k_1, k_2, \cdots, k_T), \\ 0 & \text{if} \quad x^n \ne \varphi_n(k_1, k_2, \cdots, k_T). \end{cases} \tag{17}$$

On the other hand, any stochastic encoder can be represented by the concatenation of a deterministic encoder $\widetilde{\varphi}$ and a channel $\mathbf{U}$ with alphabets $\widetilde{\mathcal{X}} \to \mathcal{X}$, which are defined by

$$\widetilde{\varphi}_n : (k_1, k_2, \cdots, k_T) \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_T \longmapsto \widetilde{x}^n_{k_1,k_2,\cdots,k_T} \in \widetilde{\mathcal{X}}^n, \tag{18}$$

$$U^n(x^n|\widetilde{x}^n) \equiv \begin{cases} Q_{k_1,k_2,\cdots,k_T}(x^n) & \text{if} \quad \widetilde{x}^n = \widetilde{\varphi}_n(k_1, k_2, \cdots, k_T), \\ \widetilde{U}^n(x^n|\widetilde{x}^n) & \text{if} \quad \widetilde{x}^n \ne \widetilde{\varphi}_n(k_1, k_2, \cdots, k_T) \\ & \qquad \text{for any } (k_1, k_2, \cdots, k_T), \end{cases} \tag{19}$$

where $\widetilde{U}^n(x^n|\widetilde{x}^n)$ is an arbitrary channel. Then, $Q_{k_1,k_2,\cdots,k_T}(x^n)$ can be described as

$$Q_{k_1,k_2,\cdots,k_T}(x^n) = \sum_{\widetilde{x}^n} U^n(x^n|\widetilde{x}^n) \delta_{\widetilde{\varphi}_n(k_1,k_2,\cdots,k_T)}(\widetilde{x}^n). \tag{20}$$

We use both descriptions interchangeably.

Bob decodes a tuple of messages $(\widehat{K}_1, \widehat{K}_2, \cdots, \widehat{K}_T)$ by a decoder ψ_n from the channel output Y^n. Let $\mathcal{D}_{k_1,k_2,\cdots,k_T}$ be the decoding region of $(k_1, k_2, \cdots, k_T) \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_T$ such that $\{\mathcal{D}_{k_1,k_2,\cdots,k_T}\}$ are mutually disjoint. Then $\widehat{K}_1 = k_1, \widehat{K}_2 = k_2, \cdots, \widehat{K}_T = k_T$ are decoded if $Y^n \in \mathcal{D}_{k_1,k_2,\cdots,k_T}$. Equivalently, we can define the decoder

[1]This example is given in [5, Remark 6.3.3].

ψ_n^t from ψ_n for each message K_t, $t = 1, 2, \cdots, T$, such that the decoding region $\mathcal{D}_k^t$ of $k \in \mathcal{K}_t$ is given by

$$\mathcal{D}_k^t \equiv \big\{ y^n \in \mathcal{Y}^n \;\big|\; y^n \in \mathcal{D}_{k_1,\cdots,k_{t-1},k,k_{t+1},\cdots,k_T} \text{ for some } k_1, \cdots, k_{t-1}, k_{t+1}, \cdots, k_T \big\}. \tag{21}$$

Then for each t, $\mathcal{D}_1^t, \mathcal{D}_2^t, \cdots, \mathcal{D}_{M_t}^t$ are mutually disjoint, and $\widehat{K}_t = k$ is decoded if $Y^n \in \mathcal{D}_k^t$. In the case that $\widehat{K}_t \neq K_t$ or $Y^n \notin \cup_{k=1}^{M_t} \mathcal{D}_k^t$, a decoding error occurs for the t-th message K_t. Note that for each $(k_1, k_2, \cdots, k_T) \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_T$,

$$\mathcal{D}_{k_1,k_2,\cdots,k_T} = \mathcal{D}_{k_1}^1 \cap \mathcal{D}_{k_2}^2 \cap \cdots \cap \mathcal{D}_{k_T}^T. \tag{22}$$

The above code is denoted by $\mathcal{C}_n(\{M_1, \cdots, M_T\}, \varphi_n, \psi_n)$, or $\mathcal{C}_n$ for short, and we evaluate the performance of the code in the following three viewpoints.

(a) Coding rate of each message K_t:

$$\frac{1}{n} \log M_t.$$

(b) Average decoding error probability of each message K_t:

$$\begin{aligned} \varepsilon_n^t(\mathcal{C}_n) &\equiv \frac{1}{M_t} \sum_{k=1}^{M_t} \Pr\big\{ Y_n \notin \mathcal{D}_k^t \;\big|\; K_t = k \big\} \\ &= \frac{1}{M_t} \sum_{k=1}^{M_t} Q_k^t W^n(\overline{\mathcal{D}_k^t}). \end{aligned} \tag{23}$$

Here $\overline{\mathcal{D}_k^t}$ is the complement set of $\mathcal{D}_k^t$ and Q_k^t is the probability distribution on $\mathcal{X}^n$ defined by

$$\begin{aligned} Q_k^t(x^n) &\equiv \Pr\big\{ X_n = x^n \;\big|\; K_t = k \big\} \\ &= \frac{1}{L_t} \sum_{(k_1,\cdots,k_{t-1},k_{t+1}\cdots,k_T) \in \mathcal{L}_t} Q_{k_1,\cdots,k_{t-1},k,k_{t+1}\cdots,k_T}(x^n), \end{aligned} \tag{24}$$

where

$$\mathcal{L}_t \equiv \mathcal{K}_1 \times \cdots \times \mathcal{K}_{t-1} \times \mathcal{K}_{t+1} \times \cdots \times \mathcal{K}_T, \tag{25}$$

$$L_t \equiv |\mathcal{L}_t| = \frac{\prod_{t=1}^T M_t}{M_t}. \tag{26}$$

In the case of deterministic encoders, we use the description in (16). The probability distributions of the outputs Y^n and Z^n of the main and wiretap channels for the message $K_t = k$ are given by $Q_k^t W^n(y^n)$ and $Q_k^t V^n(z^n)$, respectively.

(c) Security measures:

$$\begin{aligned} I_n^t(\mathcal{C}_n) &\equiv \frac{1}{n} I(K_t; Z^n) \\ &= \frac{1}{n} \frac{1}{M_t} \sum_{k=1}^{M_t} D(Q_k^t V^n || P_{Z^n}) \\ &= \frac{1}{n} \frac{1}{M_t} \sum_{k=1}^{M_t} D\left(Q_k^t V^n \;\middle\|\; \frac{1}{M_t} \sum_{k=1}^{M_t} Q_k^t V^n \right), \end{aligned} \tag{27}$$

$$\begin{aligned} d_n^t(\mathcal{C}_n) &\equiv \frac{1}{M_t(M_t-1)} \sum_{k=1}^{M_t} \sum_{k'=1(k' \neq k)}^{M_t} \big\| Q_{k'}^t V^n - Q_k^t V^n \big\|_1 \\ &= \frac{1}{M_t(M_t-1)} \sum_{k=1}^{M_t} \sum_{k'=1(k' \neq k)}^{M_t} \\ &\qquad \sum_{z^n} |Q_{k'}^t V^n(z^n) - Q_k^t V^n(z^n)|, \end{aligned} \tag{28}$$

where $D(\cdot\|\cdot)$ and $\|\cdot\|_1$ stand for a relative entropy and a variational distance, respectively.

Note that if $I_n^t(\mathcal{C}_n)$ is sufficiently small, then the message K_t and the output Z^n are almost independent, and hence, Eve cannot obtain almost any information about K_t from Z^n. On the other hand, $d_n^t(\mathcal{C}_n)$ is the security measure based on the variational distance. If $d_n^t(\mathcal{C}_n)$ is sufficiently small, then the difference between the output probability distributions $Q_k^t V^n$ and $Q_{k'}^t V^n$ is almost zero on the average for all $k, k' \in \mathcal{K}_t$. This also means that Eve cannot obtain almost any information about K_t from Z^n on the average.

Remark 3: Although we first use the average criteria for the error probability (23) and the security measure (28) following [9], the same results hold even if we employ the maximum criteria for the error probability and the variational distance. See Definitions 7–9, Theorems 4 and 5, and Section IV-D.

Now we define the achievable rates R_t, $t = 1, 2, \cdots, T$, for the multiplex coding as follows.

Definition 3: If there exists a sequence of code $\mathcal{C}_n$ that satisfies (29)–(32), then a rate-tuple $(R_1, R_2, \cdots, R_T)$ is called achievable for channels $(\mathbf{W}, \mathbf{V})$ in the sense of the security measure $I_n^t(\mathcal{C}_n)$. Furthermore, if there exists a sequence of code $\mathcal{C}_n$ that satisfies (29)–(31) and (33), then a rate-tuple $(R_1, R_2, \cdots, R_T)$ is called achievable for channels $(\mathbf{W}, \mathbf{V})$ in the sense of the security measure $d_n^t(\mathcal{C}_n)$.

$$\liminf_{n\to\infty} \frac{1}{n} \log \left(\prod_{t'=1}^T M_{t'} \right) \geq R_{\text{total}}, \tag{29}$$

$$\limsup_{n\to\infty} \left[\frac{1}{n} \log \left(\prod_{t'=1}^T M_{t'} \right) - \frac{1}{n} \log M_t \right] \leq R_{\text{total}} - R_t, \quad t = 1, 2, \cdots, T, \tag{30}$$

$$\lim_{n\to\infty} \varepsilon_n^t(\mathcal{C}_n) = 0, \quad t = 1, 2, \cdots, T, \tag{31}$$

$$\lim_{n\to\infty} I_n^t(\mathcal{C}_n) = 0, \quad t = 1, 2, \cdots, T, \tag{32}$$

$$\lim_{n\to\infty} d_n^t(\mathcal{C}_n) = 0, \quad t = 1, 2, \cdots, T, \tag{33}$$

where the total rate R_{total} is defined as

$$R_{\text{total}} = \sum_{t=1}^{T} R_t. \tag{34}$$

Remark 4: Note from (29) and (30) that if $(R_1, R_2, \cdots, R_T)$ is achievable, then it satisfies

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n^t \geq R_t, \quad t = 1, 2, \cdots, T, \tag{35}$$

because

$$\begin{aligned} R_t &\leq R_{\text{total}} - \limsup_{n\to\infty} \left[\frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) - \frac{1}{n} \log M_t \right] \\ &\leq \liminf_{n\to\infty} \frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) \\ &\quad + \liminf_{n\to\infty} \left[-\frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) + \frac{1}{n} \log M_t \right] \\ &\leq \liminf_{n\to\infty} \left[\frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) \right. \\ &\quad \left. -\frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) + \frac{1}{n} \log M_t \right] \\ &= \liminf_{n\to\infty} \frac{1}{n} \log M_t. \end{aligned} \tag{36}$$

Therefore, if $(R_1, R_2, \cdots, R_T)$ is achievable, each K_t can be transmitted securely with at least the rate R_t. However, for any rate-tuple $(R_1, R_2, \cdots, R_T)$, there exists a sequence of code $\{\mathcal{C}_n\}$ that does not satisfy both equalities of (29) and (35). Such a case occurs if $\{\mathcal{C}_n\}$ satisfies

$$\liminf_{n\to\infty} \frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) > \sum_{t'=1}^{T} \liminf_{n\to\infty} \frac{1}{n} \log M_{t'}. \tag{37}$$

In order to avoid this inconvenience, (30) is used instead of (35).

Remark 5: Although (32) and (33) ensure the perfect secrecy of each message K_t against Eve observing $\mathbf{Z}$, she can get some information about the combination of $(K_1, K_2, \cdots, K_T)$ because the entire $(K_1, K_2, \cdots, K_T)$ is not independent of $\mathbf{Z}$. But, since $K_1, K_2, \cdots, K_T$ are mutually independent, the leaked information has no meaning. It is worth noting that $K_1, K_2, \cdots, K_T$ are not mutually independent when $\mathbf{Z}$ is given. Hence, if Eve gets a message K_t by a method other than the output $\mathbf{Z}$ of the wiretap channel, she can also get some information about other messages $K_{t'}$, $t' \neq t$, from K_t and $\mathbf{Z}$. If Alice and Bob want to prevent the possibility of such attack, they must use the ordinary, i.e. non-multiplex, coding for wiretap channels.

Definition 4: Let $\mathcal{R}^I_{\text{det}}(\mathbf{W}, \mathbf{V}, T)$, $\mathcal{R}^d_{\text{det}}(\mathbf{W}, \mathbf{V}, T)$, $\mathcal{R}^I_{\text{sto}}(\mathbf{W}, \mathbf{V}, T)$, and $\mathcal{R}^d_{\text{sto}}(\mathbf{W}, \mathbf{V}, T)$ be the closures of achievable rate-tuples $(R_1, R_2, \cdots, R_T)$ for the main and wiretap channels $(\mathbf{W}, \mathbf{V})$. The subscript "det" denotes the case that only deterministic encoders can be used while "sto" means that stochastic encoders including deterministic encoders can be used. Furthermore, the superscripts "I" and "d" stand for the cases that the security is measured by $I_n^t(\mathcal{C}_n)$ and $d_n^t(\mathcal{C}_n)$, respectively.

From the definition, it holds obviously that for any $(\mathbf{W}, \mathbf{V})$ and any T,

$$\mathcal{R}^I_{\text{det}}(\mathbf{W}, \mathbf{V}, T) \subseteq \mathcal{R}^I_{\text{sto}}(\mathbf{W}, \mathbf{V}, T), \tag{38}$$

$$\mathcal{R}^d_{\text{det}}(\mathbf{W}, \mathbf{V}, T) \subseteq \mathcal{R}^d_{\text{sto}}(\mathbf{W}, \mathbf{V}, T). \tag{39}$$

Since the multiplex coding of plural messages is treated, we usually assume in this paper that $T \geq 2$. But, note that the case of $T = 1$ corresponds to the ordinary coding for wiretap channels.

In order to evaluate the above achievable rate regions, we first define four regions $\mathcal{R}^i_1(\mathbf{W}, \mathbf{V}, T)$, $\mathcal{R}^o_1(\mathbf{W}, \mathbf{V}, T)$, $\mathcal{R}^i_2(\mathbf{W}, \mathbf{V}, T)$, and $\mathcal{R}^o_2(\mathbf{W}, \mathbf{V}, T)$ as follows.

Definition 5:

$$\begin{aligned} &\mathcal{R}^i_1(\mathbf{W}, \mathbf{V}, T) \\ &\quad \equiv \big\{ (R_1, R_2, \cdots, R_T) \;\big|\; \text{There exists an input probability} \\ &\qquad \text{distribution } P_{\mathbf{X}} \text{ that satisfies (42) and (43)} \big\}, \end{aligned} \tag{40}$$

$$\begin{aligned} &\mathcal{R}^o_1(\mathbf{W}, \mathbf{V}, T) \\ &\quad \equiv \big\{ (R_1, R_2, \cdots, R_T) \;\big|\; \text{There exists an input probability} \\ &\qquad \text{distribution } P_{\mathbf{X}} \text{ that satisfies (42) and (44)} \big\}, \end{aligned} \tag{41}$$

$$R_{\text{total}} \leq \underline{I}(P_{\mathbf{X}}, \mathbf{W}), \tag{42}$$

$$R_{\text{total}} - R_t \geq \overline{I}(P_{\mathbf{X}}, \mathbf{V}), \quad t = 1, 2, \cdots, T, \tag{43}$$

$$R_{\text{total}} - R_t \geq S_{\mathbf{X}}(\mathbf{V}), \quad t = 1, 2, \cdots, T. \tag{44}$$

Definition 6:

$$\begin{aligned} &\mathcal{R}^i_2(\mathbf{W}, \mathbf{V}, T) \\ &\quad \equiv \big\{ (R_1, R_2, \cdots, R_T) \;\big|\; \text{There exists an input probability} \\ &\qquad \text{distribution } P_{\widetilde{\mathbf{X}}} \text{ and a test channel } \mathbf{U} \text{ with} \\ &\qquad \text{alphabets } \widetilde{\mathcal{X}} \to \mathcal{X} \text{ that satisfy (47) and (48)} \big\}, \end{aligned} \tag{45}$$

$$\begin{aligned} &\mathcal{R}^o_2(\mathbf{W}, \mathbf{V}, T) \\ &\quad \equiv \big\{ (R_1, R_2, \cdots, R_T) \;\big|\; \text{There exists an input probability} \\ &\qquad \text{distribution } P_{\widetilde{\mathbf{X}}} \text{ and a test channel } \mathbf{U} \text{ with} \\ &\qquad \text{alphabets } \widetilde{\mathcal{X}} \to \mathcal{X} \text{ that satisfy (47) and (49)} \big\}, \end{aligned} \tag{46}$$

$$R_{\text{total}} \leq \underline{I}(P_{\widetilde{\mathbf{X}}}, \mathbf{UW}), \tag{47}$$

$$R_{\text{total}} - R_t \geq \overline{I}(P_{\widetilde{\mathbf{X}}}, \mathbf{UV}), \quad t = 1, 2, \cdots, T, \tag{48}$$

$$R_{\text{total}} - R_t \geq S_{\widetilde{\mathbf{X}}}(\mathbf{UV}), \quad t = 1, 2, \cdots, T. \tag{49}$$

where $\widetilde{\mathcal{X}}$ is an arbitrary finite alphabet, and $\mathbf{UW}$ with $\widetilde{\mathcal{X}} \to \mathcal{Y}$ and $\mathbf{UV}$ with $\widetilde{\mathcal{X}} \to \mathcal{Z}$ are the cascade channels of ($\mathbf{U}$ and $\mathbf{W}$) and ($\mathbf{U}$ and $\mathbf{V}$), respectively.

Note from (12) that $\mathcal{R}^o_1(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}^i_1(\mathbf{W}, \mathbf{V}, T)$ and $\mathcal{R}^o_2(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}^i_2(\mathbf{W}, \mathbf{V}, T)$. For these rate regions, the following theorems hold.

Theorem 2: For any channels $\mathbf{W}$, $\mathbf{V}$, and $T \geq 2$, $\mathcal{R}^I_{\text{det}}(\mathbf{W}, \mathbf{V}, T)$ and $\mathcal{R}^d_{\text{det}}(\mathbf{W}, \mathbf{V}, T)$ satisfy

$$\mathcal{R}^I_{\text{det}}(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}^i_1(\mathbf{W}, \mathbf{V}, T), \tag{50}$$

$$\mathcal{R}^o_1(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}^d_{\text{det}}(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}^i_1(\mathbf{W}, \mathbf{V}, T), \tag{51}$$

respectively. Furthermore, if $\mathbf{V}$ satisfies (13) for any $P_{\mathbf{X}}$, then it holds that

$$\mathcal{R}^d_{\text{det}}(\mathbf{W},\mathbf{V},T)=\mathcal{R}^i_1(\mathbf{W},\mathbf{V},T)=\mathcal{R}^o_1(\mathbf{W},\mathbf{V},T). \quad (52)$$

Theorem 3: For any channels $\mathbf{W}$, $\mathbf{V}$, and $T\geq 2$, $\mathcal{R}^I_{\text{sto}}(\mathbf{W},\mathbf{V},T)$ and $\mathcal{R}^d_{\text{sto}}(\mathbf{W},\mathbf{V},T)$ satisfy

$$\mathcal{R}^I_{\text{sto}}(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^i_2(\mathbf{W},\mathbf{V},T), \quad (53)$$

$$\mathcal{R}^o_2(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^d_{\text{sto}}(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^i_2(\mathbf{W},\mathbf{V},T), \quad (54)$$

respectively.

Remark 6: We note from [9, Theorem 5 and the proof of Lemma 5] that in the case of $T=1$ the secrecy capacity C_S is given by

$$\begin{aligned}C_S&=\sup_{P_{\widetilde{\mathbf{X}}},\mathbf{U}:\,\overline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UV})=0}\underline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UW})\\&=\sup_{P_{\widetilde{\mathbf{X}}},\mathbf{U}}\left[\underline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UW})-\overline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UV})\right]\end{aligned} \quad (55)$$

in both cases of the security measures $I^t_n(\mathcal{C}_n)$ and $d^t_n(\mathcal{C}_n)$. On the other hand, it holds from Definition 6 that for $T=1$, $R_1\leq\underline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UW})$, $0=\overline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UV})=S_{\widetilde{\mathbf{X}}}(\mathbf{UV})$. Hence, it holds for $T=1$ that

$$\begin{aligned}\mathcal{R}^I_{\text{sto}}(\mathbf{W},\mathbf{V},1)&=\mathcal{R}^d_{\text{sto}}(\mathbf{W},\mathbf{V},1)=\mathcal{R}^i_2(\mathbf{W},\mathbf{V},1)\\&=\mathcal{R}^o_2(\mathbf{W},\mathbf{V},1)=[0,C_S].\end{aligned} \quad (56)$$

Remark 7: From the inner bound $\mathcal{R}^i_1(\mathbf{W},\mathbf{V},T)$ in Theorem 2, i.e., (42) and (43), we note that if R_t satisfies

$$R_t\leq\sup_{P_{\mathbf{X}}}[\underline{I}(P_{\mathbf{X}},\mathbf{W})-\overline{I}(P_{\mathbf{X}},\mathbf{V})], \quad (57)$$

then R_t can be achieved by setting other $R_{t'}$ appropriately. Similarly, from the inner bound $\mathcal{R}^i_2(\mathbf{W},\mathbf{V},T)$ in Theorem 3, i.e., (47) and (48), R_t can be achieved by setting other $R_{t'}$ appropriately if it satisfies

$$R_t\leq\sup_{P_{\widetilde{\mathbf{X}}},\mathbf{U}}[\underline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UW})-\overline{I}(P_{\widetilde{\mathbf{X}}},\mathbf{UV})]=C_S. \quad (58)$$

This means that at least one R_t can be increased to the secrecy capacity.

Remark 8: Let $P^*_{\mathbf{X}}$ be the input probability distribution that can attain the channel capacity $C=\sup_{P_{\mathbf{X}}}\underline{I}(P_{\mathbf{X}},\mathbf{W})$, i.e. $\underline{I}(P^*_{\mathbf{X}},\mathbf{W})=C$. Then, from Theorem 2 and using this $P^*_{\mathbf{X}}$ in (40), a rate-tuple $(R_1,R_2,\cdots,R_T)$ is achievable if it satisfies

$$R_{\text{total}}=C, \quad (59)$$

$$R_{\text{total}}-R_t\geq\overline{I}(P^*_{\mathbf{X}},\mathbf{V}). \quad (60)$$

Note that in the case of $R_1=R_2=\cdots=R_T$, (60) holds for T satisfying

$$T\geq\left\lceil\frac{\underline{I}(P^*_{\mathbf{X}},\mathbf{W})}{\underline{I}(P^*_{\mathbf{X}},\mathbf{W})-\overline{I}(P^*_{\mathbf{X}},\mathbf{V})}\right\rceil\geq\left\lceil\frac{C}{C_S}\right\rceil. \quad (61)$$

Therefore, the channel capacity can be attained as the total rate of the multiplex coding with an appropriate T, and each message K_t can be individually protected perfectly against Eve.

Following [9], we have adopted the average criteria for the error probability $\varepsilon^t_n(\mathcal{C}_n)$ and the security measure $d^t_n(\mathcal{C}_n)$ in Theorems 2 and 3. But, even if the average error probability $\varepsilon^t_n(\mathcal{C}_n)$ and average variational distance $d^t_n(\mathcal{C}_n)$ is small, it might occur that the error probability $\Pr\{Y_n\notin\mathcal{D}^t_k\mid K_t=k\}$ and/or the variational distance $\|Q^t_{k'}V^n-Q^t_kV^n\|_1$ become large for some $k,k'\in\mathcal{K}_t$. To overcome this defect, we next consider the maximum criteria as follows.

Definition 7: For a code $\mathcal{C}_n$, let

$$\varepsilon'^t_n(\mathcal{C}_n)\equiv\max_{1\leq k\leq M_t}Q^t_kW^n(\overline{\mathcal{D}^t_k}), \quad (62)$$

$$d'^t_n(\mathcal{C}_n)\equiv\max_{1\leq k,k'\leq M_t}\left\|Q^t_{k'}V^n-Q^t_kV^n\right\|_1, \quad (63)$$

be the maximum error probability and the worst security measure corresponding to (23) and (28), respectively.

Definition 8: If there exists a sequence of code $\mathcal{C}_n$ that satisfies (29), (30), (64), and (65), then a rate-tuple $(R_1,R_2,\cdots,R_T)$ is called achievable for the channels $(\mathbf{W},\mathbf{V})$ in the sense of the maximum criteria.

$$\lim_{n\to\infty}\varepsilon'^t_n(\mathcal{C}_n)=0, \quad (64)$$

$$\lim_{n\to\infty}d'^t_n(\mathcal{C}_n)=0. \quad (65)$$

Definition 9: We define the regions $\mathcal{R}^{\varepsilon',d'}_{\text{det}}(\mathbf{W},\mathbf{V},T)$ and $\mathcal{R}^{\varepsilon',d'}_{\text{sto}}(\mathbf{W},\mathbf{V},T)$ by the closures of achievable rate-tuples $(R_1,R_2,\cdots,R_T)$ for the channels $(\mathbf{W},\mathbf{V})$ in the sense of the maximum criteria, where the subscript "det" denotes the case that only deterministic encoders can be used, and "sto" means that stochastic encoders including deterministic encoders can be used.

Theorem 4: For any channels $\mathbf{W}$, $\mathbf{V}$, and $T\geq 2$, $\mathcal{R}^{\varepsilon',d'}_{\text{det}}(\mathbf{W},\mathbf{V},T)$ satisfies

$$\mathcal{R}^o_1(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^{\varepsilon',d'}_{\text{det}}(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^i_1(\mathbf{W},\mathbf{V},T). \quad (66)$$

Furthermore, if $\mathbf{V}$ satisfies (13) for any $P_{\mathbf{X}}$, then it holds that

$$\mathcal{R}^{\varepsilon',d'}_{\text{det}}(\mathbf{W},\mathbf{V},T)=\mathcal{R}^i_1(\mathbf{W},\mathbf{V},T)=\mathcal{R}^o_1(\mathbf{W},\mathbf{V},T). \quad (67)$$

Theorem 5: For any channels $\mathbf{W}$, $\mathbf{V}$, and $T\geq 2$, $\mathcal{R}^{\varepsilon',d'}_{\text{sto}}(\mathbf{W},\mathbf{V},T)$ satisfies

$$\mathcal{R}^o_2(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^{\varepsilon',d'}_{\text{sto}}(\mathbf{W},\mathbf{V},T)\supseteq\mathcal{R}^i_2(\mathbf{W},\mathbf{V},T). \quad (68)$$

We note that Remark 6 also holds for these maximum criteria.

IV. Proofs

A. Direct Part of Theorem 2

The direct part, i.e. (50) and the right inclusion of (51), can be proved in the same way as [9, the proof of Theorem 3], which uses the coding scheme introduced in [8].

In a code $\mathcal{C}_n$, the total number of codewords is given by $\prod_{t=1}^T M_t$. We generate every codeword independently with probability P_{X^n}. Then, let $x^n_{k_1,k_2,\cdots,k_T}$ be the codeword that corresponds to messages $K_t=k_t$, $t=1,2,\cdots,T$. The decoding regions $\mathcal{D}_{k_1,k_2,\cdots,k_T}$ for messages $K_t=k_t$, $t=1,2,\cdots,T$, are defined by

$$\mathcal{D}_{k_1,k_2,\cdots,k_T}\equiv\mathcal{A}(x^n_{k_1,k_2,\cdots,k_T})\Big\backslash\bigcup_{\substack{(k'_1,k'_2,\cdots,k'_T)\\\neq(k_1,k_2,\cdots,k_T)}}\mathcal{A}(x^n_{k'_1,k'_2,\cdots,k'_T}), \quad (69)$$

where $\mathcal{A}(x^n)$ is defined for a given real number a, which is determined later, as follows.

$$\mathcal{A}(x^n) \equiv \left\{ y^n \in \mathcal{Y}^n \;\middle|\; \frac{W^n(y^n|x^n)}{P_{Y^n}(y^n)} > 2^{an} \right\}. \tag{70}$$

We now evaluate the performance for the above random code ensemble. For the case of non-multiplex coding, Hayashi [9, Section IV] proved the coding theorem for wiretap channels by using a dummy message with size L to keep a true message with size M secret against Eve based on channel resolvability coding. In the case of multiplex coding, for each message K_t, $t = 1, 2, \cdots, T$, the other messages can be considered as a dummy message to keep the message K_t secret against Eve. Hence, we don't need to use a dummy message explicitly. For each t, the above random code ensemble coincides with Hayashi's random code ensemble with the message size $M = M_t$ and the dummy size $L = L_t$ defined in (26).

To evaluate the security measure $d_n^t(\mathcal{C}_n)$, we consider the following quantity relevant to channel resolvability coding

$$\widehat{d}_n^t(\mathcal{C}_n) \equiv \frac{1}{M_t} \sum_{k=1}^{M_t} \left\| Q_k^t V^n - P_{X^n} V^n \right\|_1, \quad t = 1, 2, \cdots, T, \tag{71}$$

which is a measure of the approximation of $P_{X^n}V$ by $Q_k^t V$ on the average. Note that $\widehat{d}_n^t(\mathcal{C}_n)$ is related to $d_n^t(\mathcal{C}_n)$ by the triangle inequality:

$$\begin{aligned} d_n^t(\mathcal{C}_n) &= \frac{1}{M_t(M_t-1)} \sum_{k=1}^{M_t} \sum_{k'=1(k'\neq k)}^{M_t} \left\| Q_{k'}^t V^n - Q_k^t V^n \right\|_1 \\ &\le \frac{1}{M_t(M_t-1)} \sum_{k=1}^{M_t} \sum_{k'=1(k'\neq k)}^{M_t} \left\{ \left\| Q_k^t V^n - P_{X^n} V^n \right\|_1 \right. \\ &\qquad \left. + \left\| Q_{k'}^t V^n - P_{X^n} V^n \right\|_1 \right\} \\ &= 2\widehat{d}_n^t(\mathcal{C}_n). \end{aligned} \tag{72}$$

Then the following Lemma holds from [9, the proof of Theorem 3 and Lemma 2].

Lemma 1: The above random code ensemble satisfies that for any real numbers a, b and for $t = 1, 2, \cdots, T$,

$$\begin{aligned} \mathrm{E}\varepsilon_n^t(\mathcal{C}_n) &\le \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} < a \right\} + L_t \cdot M_t \cdot 2^{-an} \\ &= \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} < a \right\} \\ &\quad + \left(\prod_{t=1}^{T} M_t \right) \cdot 2^{-an}, \end{aligned} \tag{73}$$

$$\mathrm{E}\widehat{d}_n^t(\mathcal{C}_n) \le 2 \Pr\left\{ \frac{1}{n} \log \frac{V^n(Z^n|X^n)}{P_{Z^n}(Z^n)} > b \right\} + \sqrt{\frac{2^{bn}}{L_t}}, \tag{74}$$

$$\mathrm{E}I_n^t(\mathcal{C}_n) \le \frac{1}{n}\eta(\delta_n) + \delta_n \cdot \log|\mathcal{Z}| + \frac{2^{bn}}{L_t}, \tag{75}$$

where E denotes the expectation over the random code ensemble, $\eta(x) = -x \log x$, and δ_n is defined by

$$\delta_n = \Pr\left\{ \frac{1}{n} \log \frac{V^n(Z^n|X^n)}{P_{Z^n}(Z^n)} > b \right\}. \tag{76}$$

It holds from Markov's inequality[2] that for each t,

$$\Pr\left\{ \left(\varepsilon_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}\varepsilon_n^t(\mathcal{C}_n) \right)^c \right\} < \frac{1}{3T}, \tag{77}$$

$$\Pr\left\{ \left(\widehat{d}_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}d_n^t(\mathcal{C}_n) \right)^c \right\} < \frac{1}{3T}, \tag{78}$$

$$\Pr\left\{ \left(I_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}I_n^t(\mathcal{C}_n) \right)^c \right\} < \frac{1}{3T}, \tag{79}$$

where $(\mathcal{G})^c$ stands for the complement event of $\mathcal{G}$. Then, we have

$$\begin{aligned} &\Pr\left\{ \bigcap_{t=1}^{T} \Big[\left(\varepsilon_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}\varepsilon_n^t(\mathcal{C}_n) \right) \right. \\ &\qquad \bigcap \left(\widehat{d}_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}d_n^t(\mathcal{C}_n) \right) \\ &\qquad \left. \bigcap \left(I_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}I_n^t(\mathcal{C}_n) \right) \Big] \right\} \\ &\ge 1 - \sum_{t=1}^{T} \Big[\Pr\left\{ \left(\varepsilon_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}\varepsilon_n^t(\mathcal{C}_n) \right)^c \right\} \\ &\qquad + \Pr\left\{ \left(\widehat{d}_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}d_n^t(\mathcal{C}_n) \right)^c \right\} \\ &\qquad + \Pr\left\{ \left(I_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}I_n^t(\mathcal{C}_n) \right)^c \right\} \Big] \\ &> 0 \end{aligned} \tag{80}$$

Hence, among the random code ensemble, there exists a code satisfying all of (81)–(83).

$$\varepsilon_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}\varepsilon_n^t(\mathcal{C}_n), \quad t = 1, 2, \cdots, T, \tag{81}$$

$$\widehat{d}_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}d_n^t(\mathcal{C}_n), \quad t = 1, 2, \cdots, T, \tag{82}$$

$$I_n^t(\mathcal{C}_n) \le 3T \cdot \mathrm{E}I_n^t(\mathcal{C}_n), \quad t = 1, 2, \cdots, T. \tag{83}$$

Now we show by selecting parameters M_t, a, and b adequately that for an arbitrarily given $P_{\mathbf{X}}$ and $\gamma > 0$, a rate-tuple $(R_1, R_2, \cdots, R_T)$ is achievable if it satisfies that

$$R_{\text{total}} \le \underline{I}(P_{\mathbf{X}}, \mathbf{W}) - \gamma, \tag{84}$$

$$R_{\text{total}} - R_t \ge \overline{I}(P_{\mathbf{X}}, \mathbf{V}) + \gamma, \quad t = 1, 2, \cdots, T. \tag{85}$$

Setting $M_t = 2^{nR_t}$, we have that

$$\prod_{t=1}^{T} M_t = 2^{nR_{\text{total}}} \le 2^{n(\underline{I}(P_{\mathbf{X}}, \mathbf{W}) - \gamma)}, \tag{86}$$

$$L_t = 2^{n(R_{\text{total}} - R_t)} \ge 2^{n(\overline{I}(P_{\mathbf{X}}, \mathbf{V}) + \gamma)}. \tag{87}$$

By setting $a = \underline{I}(P_{\mathbf{X}}, \mathbf{W}) - \gamma/2$, and $b = \overline{I}(P_{\mathbf{X}}, \mathbf{V}) + \gamma/2$, we obtain that

$$\left(\prod_{t=1}^{T} M_t \right) \cdot 2^{-an} \le 2^{-n\gamma/2}, \tag{88}$$

$$\frac{2^{bn}}{L_t} \le 2^{-n\gamma/2}. \tag{89}$$

Hence, from (72)–(75), (81)–(83), and Definition 1, it holds that

$$\lim_{n\to\infty} \varepsilon_n^t(\mathcal{C}_n) = 0, \quad t = 1, 2, \cdots, T, \tag{90}$$

$$\lim_{n\to\infty} d_n^t(\mathcal{C}_n) \le \lim_{n\to\infty} 2\widehat{d}_n^t(\mathcal{C}_n) = 0, \quad t = 1, 2, \cdots, T, \tag{91}$$

$$\lim_{n\to\infty} I_n^t(\mathcal{C}_n) = 0, \quad t = 1, 2, \cdots, T, \tag{92}$$

[2] $\Pr\{G > a\} < \frac{\mathrm{E}G}{a}$ for any non-negative random variable G and any positive constant a.

which means that the rate-tuple $(R_1, R_2, \cdots, R_T)$ is achievable if it satisfies (84) and (85). Finally, since the above argument holds for any $\gamma > 0$, any rate-tuple in $\mathcal{R}_1^i(\mathbf{W}, \mathbf{V}, T)$ is achievable in both senses of the security measures $d_n^t(\mathcal{C}_n)$ and $I_n^t(\mathcal{C}_n)$.

B. Converse Part of Theorem 2

We prove the left inclusion of (51).

Let $(R_1, R_2, \cdots, R_T) \in \mathcal{R}_{\text{det}}^d(\mathbf{W}, \mathbf{V}, T)$. Then, there exists a sequence of code $\{\mathcal{C}_n\}$ that satisfies (29)–(31) and (33). Hence the code $\{\mathcal{C}_n\}$ satisfies that for any $\gamma > 0$ and any sufficiently large n,

$$\frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) \geq R_{\text{total}} - \gamma, \tag{93}$$

$$\frac{1}{n} \log \left(\prod_{t'=1}^{T} M_{t'} \right) - \frac{1}{n} \log M_t \leq R_{\text{total}} - R_t + \gamma, \quad t = 1, 2, \cdots, T. \tag{94}$$

Note that we can regard the code $\mathcal{C}_n$ for multiplex coding as an ordinary message transmission code to Bob for sending the tuple of messages $K_1, K_2, \cdots, K_T$. Then the average error probability of this message transmission code is given by

$$e(\mathcal{C}_n) \equiv \frac{1}{M_1 \cdots M_T} \sum_{k_1=1}^{M_1} \cdots \sum_{k_T=1}^{M_T} \Pr\left\{ Y^n \notin \mathcal{D}_{k_1, k_2, \cdots, k_T} \;\middle|\; K_1 = k_1, \cdots, K_T = k_T \right\}. \tag{95}$$

From (22) and the union bound, $e(\mathcal{C}_n)$ is related to $\varepsilon_n^t(\mathcal{C}_n)$ as follows:

$$\begin{aligned} e(\mathcal{C}_n) =& \frac{1}{M_1 \cdots M_T} \sum_{k_1=1}^{M_1} \cdots \sum_{k_T=1}^{M_T} \Pr\left\{ Y^n \in \mathcal{D}_{k_1, k_2, \cdots, k_T}^c \;\middle|\; K_1 = k_1, \cdots, K_T = k_T \right\} \\ \leq& \frac{1}{M_1 \cdots M_T} \sum_{k_1=1}^{M_1} \cdots \sum_{k_T=1}^{M_T} \sum_{t=1}^{T} \Pr\left\{ Y^n \in \mathcal{D}_{k_t}^c \;\middle|\; K_1 = k_1, \cdots, K_T = k_T \right\} \\ =& \sum_{t=1}^{T} \frac{1}{M_t} \sum_{k_t=1}^{M_t} \Pr\left\{ Y^n \in \mathcal{D}_{k_t}^c \;\middle|\; K_t = k_t \right\} \\ =& \sum_{t=1}^{T} \varepsilon_n^t(\mathcal{C}_n). \end{aligned} \tag{96}$$

Hence, let X^n be the uniform random variable that takes values in the set of codewords $\{x^n_{k_1, k_2, \cdots, k_T}\}_{k_t \in \mathcal{K}_t, t=1, \cdots, T}$, then we have from Verdú-Han's Lemma [5, Lemma 3.2.2] that for any $\gamma > 0$,

$$\begin{aligned} &\sum_{t=1}^{T} \varepsilon_n^t(\mathcal{C}_n) \\ &\geq e(\mathcal{C}_n) \\ &\geq \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \frac{1}{n} \log \left(\prod_{t=1}^{T} M_t \right) - \gamma \right\} - e^{-n\gamma} \\ &\geq \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq R_{\text{total}} - 2\gamma \right\} - e^{-n\gamma}. \end{aligned} \tag{97}$$

Hence we conclude from (3) and (31) that R_{total} must satisfy (42).

Next we prove (44) by considering the following variational distance for each t. As defined in (24), $Q_k^t(x^n)$ is the probability distribution of the input on condition that $K_t = k$. Then, noting $P_{X^n} = \frac{1}{M_t} \sum_{k=1}^{M_t} Q_k^t$, we have that

$$\begin{aligned} &\|Q_k^t V^n - P_{X^n} V^n\|_1 \\ &\quad = \left\| Q_k^t V^n - \frac{1}{M_t} \sum_{k'=1}^{M_t} Q_{k'}^t V^n \right\|_1 \\ &\quad = \frac{1}{M_t} \left\| \sum_{k'=1}^{M_t} \left(Q_k^t V^n - Q_{k'}^t V^n \right) \right\|_1 \\ &\quad = \frac{1}{M_t} \left\| \sum_{k'=1 (k' \neq k)}^{M_t} \left(Q_k^t V^n - Q_{k'}^t V^n \right) \right\|_1 \\ &\quad \leq \frac{1}{M_t} \sum_{k'=1 (k' \neq k)}^{M_t} \left\| Q_k^t V^n - Q_{k'}^t V^n \right\|_1 \\ &\quad \leq \frac{1}{M_t - 1} \sum_{k'=1 (k' \neq k)}^{M_t} \left\| Q_k^t V^n - Q_{k'}^t V^n \right\|_1. \end{aligned} \tag{98}$$

Since the average of $\|Q_k^t V^n - P_{X^n} V^n\|_1$ for $k = 1$ to M_t tends to zero asymptotically from (33) and (98), it must hold that for some sequence of $k(n) \in \mathcal{K}_t$, $n = 1, 2, \cdots$,

$$\lim_{n \to \infty} \left\| Q_{k(n)}^t V^n - P_{X^n} V^n \right\|_1 = 0. \tag{99}$$

Noting that $Q_{k(n)}^t$ is the uniform distribution over the set with $L_t = \left(\prod_{t=1}^{T} M_t \right) / M_t$ elements, $S_{\mathbf{X}}(\mathbf{V})$ must satisfy from Definition 2 and (94) that for any $\gamma > 0$,

$$S_{\mathbf{X}}(\mathbf{V}) \leq \limsup_{n \to \infty} \frac{1}{n} \log L_t \leq R_{\text{total}} - R_t + \gamma. \tag{100}$$

Hence, any $(R_1, R_2, \cdots, R_T) \in \mathcal{R}_{\text{det}}^d(\mathbf{W}, \mathbf{V}, T)$ is included in $\mathcal{R}_1^o(\mathbf{W}, \mathbf{V}, T)$.

C. Proof of Theorem 3

In the case of Theorem 3, we can use a test channel $\mathbf{U}$ with alphabets $\widetilde{\mathcal{X}} \to \mathcal{X}$ and a deterministic encoder $\widetilde{\varphi}$ defined by (18) and (19) for a stochastic encoder φ_n. Hence, Theorem 3 can be proved in the same way as Theorem 2 by considering the cascade channels $(\mathbf{UW}, \mathbf{UV})$ instead of $(\mathbf{W}, \mathbf{V})$.

D. Proof for the Maximum Criteria

We show the proof of Theorems 4 and 5. First, we show the direct part, i.e., the right inclusion of (66), in Theorem 4. We will construct a code $\mathcal{C}_n''$ for multiplex coding by applying the expurgation technique twice to the codebook $\mathcal{C}_n$ constructed in Subsection IV-A which satisfies (90) and (91). To evaluate the maximum security measure (63), we define

$$\widehat{d'}^{\,t}_n(\mathcal{C}_n) \equiv \max_{1\le k\le M_t} \left\|Q_k^t V^n - P_{X^n}V^n\right\|_1, \tag{101}$$

corresponding to (71). Then it holds that $d'^t_n(\mathcal{C}_n) \le 2\widehat{d'}^{\,t}_n(\mathcal{C}_n)$ in the same way as (72).

Let $\{\alpha_n\}_{n=1}^\infty$ be a sequence of real numbers such that

$$0<\alpha_n<1, \quad \lim_{n\to\infty}\alpha_n = 0, \quad \lim_{n\to\infty}\frac{\widehat{d}^t_n(\mathcal{C}_n)}{\alpha_n} = 0 \tag{102}$$

for any $t=1,2,\cdots,T$. For each $t=1,2,\cdots,T$, we divide the set $\mathcal{K}_t=\{1,2,\cdots,M_t\}$ of messages into two disjoint sets, $\mathcal{K}_{t,\text{good}}$ and $\mathcal{K}_{t,\text{bad}}$, with the following ratio:

$$|\mathcal{K}_{t,\text{good}}| : |\mathcal{K}_{t,\text{bad}}| = (1-\alpha_n):\alpha_n, \tag{103}$$

with respect to the security measure $\|Q_k^tV^n - P_{X^n}V^n\|_1$ so that

$$\max_{k\in\mathcal{K}_{t,\text{good}}}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1 \le \min_{k\in\mathcal{K}_{t,\text{bad}}}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1. \tag{104}$$

A codeword $x^n_{k_1,k_2,\cdots,k_T}$ in $\mathcal{C}_n$ is expurgated if $k_t\in\mathcal{K}_{t,\text{bad}}$ for some $t=1,2,\cdots,T$. We call thus obtained codebook $\mathcal{C}_n'$. Then the set of the messages in $\mathcal{C}_n'$ is $\mathcal{K}_t'\equiv\mathcal{K}_{t,\text{good}}$ with the cardinality $M_t'\equiv(1-\alpha_n)M_t$ for each $t=1,2,\cdots,T$, and hence, the coding rates of $\mathcal{C}_n'$ are asymptotically same as those of $\mathcal{C}_n$.

Similar to (24)–(26), letting

$$\mathcal{L}_t' \equiv \mathcal{K}_1'\times\cdots\times\mathcal{K}_{t-1}'\times\mathcal{K}_{t+1}'\times\cdots\times\mathcal{K}_T' \tag{105}$$

$$L_t' \equiv |\mathcal{L}_t'| = \frac{\prod_{t=1}^T M_t'}{M_t'}, \tag{106}$$

we define the probability distribution of the input X^n corresponding to the message $k\in\mathcal{K}_t'$ for the expurgated code $\mathcal{C}_n'$ by

$$Q'^t_k(x^n) \equiv \frac{1}{L_t'}\sum_{(k_1,\cdots,k_{t-1},k_{t+1}\cdots,k_T)\in\mathcal{L}_t'} Q_{k_1,\cdots,k_{t-1},k,k_{t+1},k_T}(x^n). \tag{107}$$

We also define

$$Q'^t_{k,\text{bad}}(x^n) \equiv \frac{1}{|\mathcal{L}_t\backslash\mathcal{L}_t'|}\sum_{(k_1,\cdots,k_{t-1},k_{t+1}\cdots,k_T)\in\mathcal{L}_t\backslash\mathcal{L}_t'} Q_{k_1,\cdots,k_{t-1},k,k_{t+1},k_T}(x^n). \tag{108}$$

Using (24), $L_t'=(1-\alpha_n)^{T-1}L_t$, and $|\mathcal{L}_t\backslash\mathcal{L}_t'| = \{1-(1-\alpha_n)^{T-1}\}L_t$, we have

$$\begin{aligned}
&Q_k^t(x^n)\\
&=(1-\alpha_n)^{T-1}\cdot\frac{1}{(1-\alpha_n)^{T-1}L_t}\cdot\\
&\qquad\sum_{(k_1,\cdots,k_{t-1},k_{t+1}\cdots,k_T)\in\mathcal{L}_t'} Q_{k_1,\cdots,k_{t-1},k,k_{t+1},k_T}(x^n)\\
&\quad+\{1-(1-\alpha_n)^{T-1}\}\cdot\frac{1}{\{1-(1-\alpha_n)^{T-1}\}L_t}\cdot\\
&\qquad\sum_{(k_1,\cdots,k_{t-1},k_{t+1}\cdots,k_T)\in\mathcal{L}_t\backslash\mathcal{L}_t'} Q_{k_1,\cdots,k_{t-1},k,k_{t+1},k_T}(x^n).\\
&=(1-\alpha_n)^{T-1}Q'^t_k(x^n)\\
&\qquad+\{1-(1-\alpha_n)^{T-1}\}Q'^t_{k,\text{bad}}(x^n).
\end{aligned}\tag{109}$$

Hence it holds that

$$\begin{aligned}
&\left\|Q_k^t - Q'^t_k\right\|_1\\
&=\left\|(1-\alpha_n)^{T-1}Q'^t_k+\{1-(1-\alpha_n)^{T-1}\}Q'^t_{k,\text{bad}}-Q'^t_k\right\|_1\\
&=\{1-(1-\alpha_n)^{T-1}\}\left\|Q'^t_{k,\text{bad}}-Q'^t_k\right\|_1\\
&\le 2\{1-(1-\alpha_n)^{T-1}\}.
\end{aligned}\tag{110}$$

In the same way, we have

$$\left\|Q_k^tW^n - Q'^t_kW^n\right\|_1 \le 2\{1-(1-\alpha_n)^{T-1}\}, \tag{111}$$

$$\left\|Q_k^tV^n - Q'^t_kV^n\right\|_1 \le 2\{1-(1-\alpha_n)^{T-1}\}. \tag{112}$$

Now we evaluate the maximum security measures $\widehat{d'}^{\,t}_n(\mathcal{C}_n')$, $t=1,2,\cdots,T$, for the expurgated code $\mathcal{C}_n'$. From the definition of $\mathcal{K}_t'$ and $\mathcal{K}_{t,\text{bad}}$, we have

$$\begin{aligned}
\widehat{d}^t_n(\mathcal{C}_n) &= \frac{1}{M_t}\sum_{k\in\mathcal{K}_t}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1\\
&=(1-\alpha_n)\cdot\frac{1}{(1-\alpha_n)M_t}\sum_{k\in\mathcal{K}_t'}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1\\
&\qquad+\alpha_n\cdot\frac{1}{\alpha_nM_t}\sum_{k\in\mathcal{K}_{t,\text{bad}}}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1\\
&\ge\alpha_n\cdot\frac{1}{\alpha_nM_t}\sum_{k\in\mathcal{K}_{t,\text{bad}}}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1\\
&\ge\alpha_n\cdot\max_{k\in\mathcal{K}_t'}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1,
\end{aligned}\tag{113}$$

which leads to

$$\begin{aligned}
\widehat{d'}^{\,t}_n(\mathcal{C}_n') &= \max_{k\in\mathcal{K}_t'}\left\|Q'^t_kV^n - P_{X^n}V^n\right\|_1\\
&=\max_{k\in\mathcal{K}_t'}\left\|Q_k^tV^n - P_{X^n}V^n + Q'^t_kV^n - Q_k^tV^n\right\|_1\\
&\le\max_{k\in\mathcal{K}_t'}\left\|Q_k^tV^n - P_{X^n}V^n\right\|_1\\
&\qquad+\max_{k\in\mathcal{K}_t'}\left\|Q'^t_kV^n - Q_k^tV^n\right\|_1\\
&\le\frac{\widehat{d}^t_n(\mathcal{C}_n)}{\alpha_n}+2\{1-(1-\alpha_n)^{T-1}\},
\end{aligned}\tag{114}$$

where we used the triangle inequality, (112), and (113). Thus from the above inequality and (102) we obtain

$$\lim_{n\to\infty} \widehat{d'}_n^t(\mathcal{C}'_n) = 0. \tag{115}$$

Next we evaluate the average error probabilities $\varepsilon_n^t(\mathcal{C}'_n)$, $t = 1, 2, \cdots, T$, for the expurgated code $\mathcal{C}'_n$ as follows:

$$\begin{aligned}
&\varepsilon_n^t(\mathcal{C}_n) \\
&= \frac{1}{M_t} \sum_{k\in\mathcal{K}_t} Q_k^t W^n(\overline{\mathcal{D}_k^t}), \\
&= (1-\alpha_n) \cdot \frac{1}{(1-\alpha_n)M_t} \sum_{k\in\mathcal{K}'_t} Q_k^t W^n(\overline{\mathcal{D}_k^t}) \\
&\qquad + \alpha_n \cdot \frac{1}{\alpha_n M_t} \sum_{k\in\mathcal{K}_{t,\mathrm{bad}}} Q_k^t W^n(\overline{\mathcal{D}_k^t}) \\
&\geq (1-\alpha_n) \cdot \frac{1}{M'_t} \sum_{k\in\mathcal{K}'_t} Q_k^t W^n(\overline{\mathcal{D}_k^t}) \\
&\geq (1-\alpha_n) \cdot \frac{1}{M'_t} \sum_{k\in\mathcal{K}'_t} \Big\{ {Q'}_k^t W^n(\overline{\mathcal{D}_k^t}) + Q_k^t W^n(\overline{\mathcal{D}_k^t}) \\
&\qquad\qquad - {Q'}_k^t W^n(\overline{\mathcal{D}_k^t}) \Big\} \\
&\geq (1-\alpha_n) \cdot \frac{1}{M'_t} \sum_{k\in\mathcal{K}'_t} \Big\{ {Q'}_k^t W^n(\overline{\mathcal{D}_k^t}) \\
&\qquad\qquad - \Big| Q_k^t W^n(\overline{\mathcal{D}_k^t}) - {Q'}_k^t W^n(\overline{\mathcal{D}_k^t}) \Big| \Big\} \\
&\geq (1-\alpha_n) \left[\varepsilon_n^t(\mathcal{C}'_n) - \{1-(1-\alpha_n)^{T-1}\} \right],
\end{aligned} \tag{116}$$

where we used the following inequality derived from (111),

$$\begin{aligned}
\max_{\mathcal{D}\subseteq\mathcal{Y}^n} \left| Q_k^t W^n(\mathcal{D}) - {Q'}_k^t W^n(\mathcal{D}) \right| &= \frac{1}{2} \left\| Q_k^t W^n - {Q'}_k^t W^n \right\|_1 \\
&\leq \{1-(1-\alpha_n)^{T-1}\}.
\end{aligned} \tag{117}$$

Hence it holds that

$$\lim_{n\to\infty} \varepsilon_n^t(\mathcal{C}'_n) \leq \lim_{n\to\infty} \left[\frac{\varepsilon_n^t(\mathcal{C}_n)}{1-\alpha_n} + \{1-(1-\alpha_n)^{T-1}\} \right] = 0. \tag{118}$$

As the second stage of the expurgation, we apply the same expurgation technique to the codebook $\mathcal{C}'_n$ to exclude codewords with bad decoding error probabilities. We call thus obtained code $\mathcal{C}''_n$. Then the coding rates of $\mathcal{C}''_n$ and $\mathcal{C}'_n$ are asymptotically same, and it holds that

$$\lim_{n\to\infty} {\varepsilon'}_n^t(\mathcal{C}''_n) = 0, \quad \lim_{n\to\infty} {d'}_n^t(\mathcal{C}''_n) \leq \lim_{n\to\infty} 2\widehat{d'}_n^t(\mathcal{C}''_n) = 0, \tag{119}$$

for $t = 1, 2, \cdots, T$ simultaneously. Thus we have shown the direct part of (66) in Theorem 4.

On the other hand, it is obvious from the definition that

$$\mathcal{R}_{\mathrm{det}}^{\varepsilon',d'}(\mathbf{W},\mathbf{V},T) \subseteq \mathcal{R}_{\mathrm{det}}^{d}(\mathbf{W},\mathbf{V},T), \tag{120}$$

which means, from (51), that $\mathcal{R}_{\mathrm{det}}^{\varepsilon',d'}(\mathbf{W},\mathbf{V},T) \subseteq \mathcal{R}_1^o(\mathbf{W},\mathbf{V},T)$.

V. Stationary Memoryless Wiretap Channels

In this section, we consider the case that channels $\mathbf{W}$ and $\mathbf{V}$ are stationary memoryless channels with transition probability distributions W and V, respectively. In this case, if we restrict the input source $P_{\mathbf{X}}$ and the test channel $\mathbf{U}$ to the stationary memoryless source and channel, respectively, in (40) and (45), then the spectral sup- and inf-mutual information rates are equal to the ordinary mutual information. Hence the following corollary holds from Theorems 2 and 3.

Corollary 1: If the channels $\mathbf{W}$ and $\mathbf{V}$ are stationary memoryless channels given by W and V, respectively, it holds for $T \geq 2$ that

$$\mathcal{R}_{\mathrm{det}}^{I}(\mathbf{W},\mathbf{V},T) \supseteq \mathcal{R}_1^*(W,V,T), \tag{121}$$
$$\mathcal{R}_{\mathrm{det}}^{d}(\mathbf{W},\mathbf{V},T) \supseteq \mathcal{R}_1^*(W,V,T), \tag{122}$$
$$\mathcal{R}_{\mathrm{sto}}^{I}(\mathbf{W},\mathbf{V},T) \supseteq \mathcal{R}_2^*(W,V,T), \tag{123}$$
$$\mathcal{R}_{\mathrm{sto}}^{d}(\mathbf{W},\mathbf{V},T) \supseteq \mathcal{R}_2^*(W,V,T), \tag{124}$$

where $\mathcal{R}_1^*(W,V,T)$ and $\mathcal{R}_2^*(W,V,T)$ are defined in Definition 10 below.

Definition 10:

$$\begin{aligned}
&\mathcal{R}_1^*(W,V,T) \\
&\quad \equiv \big\{ (R_1, R_2, \cdots, R_T) \;\big|\; \text{There exists an input probability distribution } P_X \text{ that satisfies (127) and (128)} \big\}.
\end{aligned} \tag{125}$$

$$\begin{aligned}
&\mathcal{R}_2^*(W,V,T) \\
&\quad \equiv \big\{ (R_1, R_2, \cdots, R_T) \;\big|\; \text{There exists an input probability distribution } P_{\widetilde{X}} \text{ and a test channel } U \text{ with} \\
&\qquad \text{alphabets } \widetilde{X} \to \mathcal{X} \text{ that satisfy (129) and (130)} \big\}.
\end{aligned} \tag{126}$$

$$R_{\mathrm{total}} \leq I(P_X, W), \tag{127}$$
$$R_{\mathrm{total}} - R_t \geq I(P_X, V), \quad t = 1, 2, \cdots, T, \tag{128}$$
$$R_{\mathrm{total}} \leq I(P_{\widetilde{X}}, UW), \tag{129}$$
$$R_{\mathrm{total}} - R_t \geq I(P_{\widetilde{X}}, UV), \quad t = 1, 2, \cdots, T, \tag{130}$$

where $\widetilde{X}$ is an auxiliary random variable over a finite alphabet $\widetilde{\mathcal{X}}$ [3], and UW with $\widetilde{X} \to \mathcal{Y}$ and UV with $\widetilde{X} \to \mathcal{Z}$ are the cascade channels of (U and W) and (U and V), respectively. The mutual information are defined as $I(P_X, W) \equiv I(X;Y)$, $I(P_X, V) \equiv I(X;Z)$, $I(P_{\widetilde{X}}, UW) \equiv I(\widetilde{X};Y)$, $I(P_{\widetilde{X}}, UV) \equiv I(\widetilde{X};Z)$ where the random variables make a Markov chain $\widetilde{X} \to X \to (Y, Z)$.

In the same way, Theorems 4 and 5 yield the following corollary with respect to the maximum criteria.

Corollary 2: If the channels $\mathbf{W}$ and $\mathbf{V}$ are stationary memoryless channels given by W and V, respectively, it holds for $T \geq 2$ that

$$\mathcal{R}_{\mathrm{det}}^{\varepsilon',d'}(\mathbf{W},\mathbf{V},T) \supseteq \mathcal{R}_1^*(W,V,T), \tag{131}$$
$$\mathcal{R}_{\mathrm{sto}}^{\varepsilon',d'}(\mathbf{W},\mathbf{V},T) \supseteq \mathcal{R}_2^*(W,V,T). \tag{132}$$

[3] $|\widetilde{\mathcal{X}}|$, the cardinality of $\widetilde{\mathcal{X}}$, can be bounded by $|\widetilde{\mathcal{X}}| \leq |\mathcal{X}| + 1$. Refer [3] for more details.

We note that δ_n defined in (76) goes to zero with an exponential order of n in the case of stationary memoryless channels. Hence, even if we use

$$\widehat{I}_n^t(\mathcal{C}_n) \equiv I(K_t; Z^n) \tag{133}$$

instead of $I_n^t(\mathcal{C}_n)$ defined in (27) as a security measure, we can easily prove that the code shown in Section IV-A also satisfies

$$\lim_{n\to\infty} \widehat{I}_n^t(\mathcal{C}_n) = 0. \tag{134}$$

Therefore, Corollary 1 holds for $\widehat{I}_n^t(\mathcal{C}_n)$ similarly. Furthermore, it is well known, e.g. refer [2], that the divergence $D(P_1||P_2)$ and the variational distance $\|P_1 - P_2\|_1$ satisfies

$$D(P_1||P_2) \geq \frac{1}{2\ln 2} \|P_1 - P_2\|_1^2. \tag{135}$$

Hence, if a rate-tuple $(R_1, R_2, \cdots, R_T)$ is achievable for $\widehat{I}_n^t(\mathcal{C}_n)$, then the rate-tuple is also achievable for $d_n^t(\mathcal{C}_n)$. This means that $\mathcal{R}_{\text{det}}^{\widehat{I}}(\mathbf{W}, \mathbf{V}, T) \subseteq \mathcal{R}_{\text{det}}^{d}(\mathbf{W}, \mathbf{V}, T)$ and $\mathcal{R}_{\text{sto}}^{\widehat{I}}(\mathbf{W}, \mathbf{V}, T) \subseteq \mathcal{R}_{\text{sto}}^{d}(\mathbf{W}, \mathbf{V}, T)$. By combining Theorems 2–4, Remark 2, and the above facts, we obtain the following corollary.

Corollary 3: If the channels $\mathbf{W}$ and $\mathbf{V}$ are stationary memoryless channels given by W and V, respectively, it holds that

$$\mathcal{R}_{\text{det}}^{d}(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}_{\text{det}}^{\widehat{I}}(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}_1^*(W, V, T) \ \text{ for } T \geq 2, \tag{136}$$

$$\mathcal{R}_{\text{sto}}^{d}(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}_{\text{sto}}^{\widehat{I}}(\mathbf{W}, \mathbf{V}, T) \supseteq \mathcal{R}_2^*(W, V, T) \ \text{ for } T \geq 2. \tag{137}$$

Furthermore, if the wiretap channel V is full-rank, then the following equalities also hold.

$$\begin{aligned}\mathcal{R}_{\text{det}}^{\varepsilon',d'}(\mathbf{W}, \mathbf{V}, T) &= \mathcal{R}_{\text{det}}^{d}(\mathbf{W}, \mathbf{V}, T) \\ &= \mathcal{R}_{\text{det}}^{\widehat{I}}(\mathbf{W}, \mathbf{V}, T) = \mathcal{R}_1^*(W, V, T) \quad \text{for } T \geq 2. \end{aligned} \tag{138}$$

We established the coding theorem for stationary memoryless full-rank wiretap channels with deterministic coding. From (127), (128), and (138), we note that if V is full-rank, every R_t must satisfy $R_t \leq I(P_X, W) - I(P_X, V)$ for some P_X. But, if V is a non-full-rank channel with $I(P_X, V) > S_{\mathbf{X}}(\mathbf{V})$, it may be possible that we can increase some R_t larger than $I(P_X, W) - I(P_X, V)$ by using small $R_{\text{total}} - R_t$. This comes from the fact that, as shown in Remark 2, the desired output distribution of $\mathbf{Z}$ can be generated by using small coding rate attaining $S_{\mathbf{X}}(\mathbf{V})$, which is less than $I(P_X, V)$. Hence, in the case of non-full-rank wiretap channels, the achievable rate region $\mathcal{R}_{\text{det}}^{d}(\mathbf{W}, \mathbf{V}, T)$ might be larger than $\mathcal{R}_1^*(W, V, T)$. But, it is an open problem to determine $S_{\mathbf{X}}(\mathbf{V})$, and hence $\mathcal{R}_{\text{det}}^{d}(\mathbf{W}, \mathbf{V}, T)$, for non-full-rank channels $\mathbf{V}$.

For any full-rank channel V, (138) holds even if W is not full-rank. However, in the case that V is a degraded version of W, i.e. in the case that $V(z|x) = \sum_y \tilde{V}(z|y) W(y|x)$ for some $\tilde{V}(z|y)$, V is not full-rank if W is not full-rank. Therefore, in the case of degraded wiretap channels, the full-rankness of W is also required for (138).

Remark 9: In the case of $T = 1$, we note from Remark 6 and Theorem 5 that for stochastic encoders, all the achievable rate regions for the security measures coincide with $\mathcal{R}_2^*(W, V, 1)$, and the maximum $R_1 \in \mathcal{R}_2^*(W, V, 1)$ is equal to the secrecy capacity C_S determined by Csiszár and Körner [3]. Hence, it holds that

$$\begin{aligned}\mathcal{R}_{\text{sto}}^{\varepsilon',d'}(\mathbf{W}, \mathbf{V}, 1) &= \mathcal{R}_{\text{sto}}^{d}(\mathbf{W}, \mathbf{V}, 1) = \mathcal{R}_{\text{sto}}^{I}(\mathbf{W}, \mathbf{V}, 1) \\ &= \mathcal{R}_{\text{sto}}^{\widehat{I}}(\mathbf{W}, \mathbf{V}, 1) = \mathcal{R}_2^*(W, V, 1) = [0, C_S], \end{aligned} \tag{139}$$

where C_S is given by

$$C_S = \max_{\widetilde{X} \to X \to (Y,Z)} \Big[I(\widetilde{X}; Y) - I(\widetilde{X}; Z) \Big]. \tag{140}$$

VI. Secure multiplex linear coding for binary symmetric wiretap channels

In this section, we show how the secure multiplex coding can be realized for the security measure $I_n^t(\mathcal{C}_n)$ by linear coding for the binary symmetric wiretap channel such that $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$, $W(0|1) = W(1|0) = p$, $V(0|1) = V(1|0) = q$, $0 \leq p < q \leq 0.5$. In this case, the channel capacity C and the secrecy capacity C_S are given by

$$C = 1 - h(p), \tag{141}$$

$$C_S = h(q) - h(p), \tag{142}$$

respectively, where $h(p)$ is the binary entropy function defined by $h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$.

For simplicity, we treat the case that $R_1 = R_2 = \cdots = R_T$ for $T = \lceil C/C_S \rceil \geq 2$. For arbitrary $\xi > 0$, define $\lambda > 0$ as $\lambda = (TC_S - C + \xi)/T$, i.e., $C - \xi = T(C_S - \lambda)$. Let each message K_t, $1 \leq t \leq T$, be a binary sequence $\boldsymbol{S}_t \equiv (S_{(t-1)\ell+1}, S_{(t-1)\ell+2}, \cdots, S_{t\ell})$ where each S_j is i.i.d. such that $\Pr\{S_j = 0\} = \Pr\{S_j = 1\} = 0.5$. We set the length ℓ as $\ell = n(C_S - \lambda)$. (For simplicity of notation, we treat $n(C_S - \lambda)$ as an integer because the difference $\lceil n(C_S - \lambda) \rceil - n(C_S - \lambda)$ can be ignored when n is sufficiently large.) Then, the coding rates are given by

$$R_t = \frac{\ell}{n} = C_S - \lambda, \tag{143}$$

$$R_{\text{total}} = \frac{T\ell}{n} = C - \xi. \tag{144}$$

Now, we define a code $\mathcal{C}_n$ by the following generator matrix G.

$$G = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_T \end{bmatrix}, \tag{145}$$

where each submatrix G_t is a binary matrix with ℓ rows and n columns. The codeword X^n is obtained by $X^n = (\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_T) G$.

We consider a random linear code ensemble such that each element of G is chosen independently at random with equal probability of being a 0 or a 1. Then, G has rank $T\ell$ with probability approaching 1 as n tends to infinity [11, Lemma 4.4], and the decoding error probability of $(\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_T)$

is bounded by $2^{-nE(R_{\rm total})}$ where $E(R_{\rm total})>0$ for $R_{\rm total}<C$ [12]. Hence, (31) is satisfied.

Next we show that the above code satisfies the security condition (32).

We first obtain the following relation.

$$\begin{aligned}
&I(\boldsymbol{S}_t;Z^n)\\
&\quad= H(Z^n)-H(Z^n|\boldsymbol{S}_t)\\
&\quad= H(Z^n)-H(Z^nX^n|\boldsymbol{S}_t)+H(X^n|Z^n\boldsymbol{S}_t)\\
&\quad= H(Z^n)-H(X^n|\boldsymbol{S}_t)-H(Z^n|X^n\boldsymbol{S}_t)+H(X^n|Z^n\boldsymbol{S}_t)\\
&\quad\overset{(a)}{=} H(Z^n)-H(X^n|\boldsymbol{S}_t)-H(Z^n|X^n)+H(X^n|Z^n\boldsymbol{S}_t)\\
&\quad\overset{(b)}{\le} n-n(T-1)(C_S-\lambda)-nh(q)+H(X^n|Z^n\boldsymbol{S}_t)\\
&\quad\overset{(c)}{=} n-n(C-\xi-C_S+\lambda)-nh(q)+H(X^n|Z^n\boldsymbol{S}_t)\\
&\quad\overset{(d)}{=} -n(\lambda-\xi)+H(X^n|Z^n\boldsymbol{S}_t),
\end{aligned}\tag{146}$$

because

(a) $\boldsymbol{S}_t\to X^n\to Z^n$ makes a Markov chain.
(b) $H(Z^n)\le n$, $H(Z^n|X^n)=nh(q)$, and $H(X^n|\boldsymbol{S}_t)=n(T-1)(C_S-\lambda)$ since $X^n=(\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_T)G$ where G has rank $T\ell$ and every $\boldsymbol{S}_t$ is independent of the others.
(c) $(T-1)(C_S-\lambda)=C-\xi-(C_S-\lambda)$.
(d) $C-C_S=1-h(q)$.

In oder to evaluate $H(X^n|Z^n\boldsymbol{S}_t)$, we split $\boldsymbol{S}_{t+1}$ and G_{t+1} as $\boldsymbol{S}_{t+1}=(\boldsymbol{S}_{t+1}^{(1)},\boldsymbol{S}_{t+1}^{(2)})$ and $G_{t+1}=\begin{bmatrix}G_{t+1}^{(1)}\\G_{t+1}^{(2)}\end{bmatrix}$, respectively, where $\boldsymbol{S}_{t+1}^{(1)}$ is the first λn bits of $\boldsymbol{S}_{t+1}$ and $G_{t+1}^{(1)}$ is the first λn rows of G_{t+1}. (When $t=T$, $t+1$ means 1.) Now consider the following submatrix $\widehat{G}_t$.

$$\widehat{G}_t=\begin{bmatrix}G_1\\\vdots\\G_{t-1}\\G_{t+1}^{(2)}\\G_{t+2}\\\vdots\\G_T\end{bmatrix},\tag{147}$$

Then, we have the relation

$$\begin{aligned}
&(\boldsymbol{S}_1,\cdots,\boldsymbol{S}_{t-1},\boldsymbol{S}_{t+1}^{(2)},\boldsymbol{S}_{t+2}\cdots,\boldsymbol{S}_T)\widehat{G}_t\\
&\qquad=(\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_T)G-\boldsymbol{S}_tG_t-\boldsymbol{S}_{t+1}^{(1)}G_{t+1}^{(1)}\\
&\qquad=X^n-\boldsymbol{S}_tG_t-\boldsymbol{S}_{t+1}^{(1)}G_{t+1}^{(1)}.
\end{aligned}\tag{148}$$

The coding rate of $\widehat{G}_t$ is given by $[(T-1)\ell-\lambda n]/n=(T-1)(C_S-\lambda)-\lambda=(C-\xi-C_S+\lambda)-\lambda=C_V-\xi$, where $C_V=C-C_S=1-h(q)$ is the channel capacity of the wiretap channel V. Since the coding rate of $\widehat{G}_t$ is less than C_V and the errors of the binary symmetric channel are linear, we can decode $X^n-\boldsymbol{S}_tG_t-\boldsymbol{S}_{t+1}^{(1)}G_{t+1}^{(1)}$ from $Z^n-\boldsymbol{S}_tG_t-\boldsymbol{S}_{t+1}^{(1)}G_{t+1}^{(1)}$ with decoding error probability $P_{t,e}\le 2^{-nE(C_V-\xi)}$. Hence, since G has rank $T\ell$, we have from Fano's inequality that for sufficiently large n,

$$\begin{aligned}
H(X^n|Z^n\boldsymbol{S}_t)&=H(X^n\boldsymbol{S}_{t+1}^{(1)}|Z^n\boldsymbol{S}_t)\\
&=H(\boldsymbol{S}_{t+1}^{(1)}|Z^n\boldsymbol{S}_t)+H(X^n|Z^n\boldsymbol{S}_t\boldsymbol{S}_{t+1}^{(1)})\\
&\le H(\boldsymbol{S}_{t+1}^{(1)})+H(X^n|Z^n\boldsymbol{S}_t\boldsymbol{S}_{t+1}^{(1)})\\
&\le n\lambda+h(P_{t,e})+(n-\ell-\lambda)P_{t,e}\\
&\le n\lambda+\xi.
\end{aligned}\tag{149}$$

Therefore, from (146) and (149), we have $I_n^t(\mathcal{C}_n)=I(\boldsymbol{S}_t;Z^n)/n\le\xi+(\xi/n)$ for every t. This means that we can realize the secure multiplex linear coding for any $R_{\rm total}<C$. In the above we considered a random linear code ensemble, but we can easily show that there exist a code satisfying all the required conditions in the ensemble.

Finally we note that in order to treat the strong security measure $\widehat{I}_n^t(\mathcal{C}_n)$ instead of the weak security measure $I_n^t(\mathcal{C}_n)$, we need more precise analysis, e.g. as shown in [13, Section 4]. We also note that if the binary wiretap channel is used for the key agreement, we can use the same technique shown in [14] to convert the weak secrecy to the strong secrecy for $(\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_T)$.

VII. Concluding Remarks

In this paper, we proved the coding theorems for multiplex wiretap channel coding by applying Hayashi's coding theorem in wiretap channels based on channel resolvability. In the case of non-multiplex wiretap channel coding, we cannot send a message securely if the coding rate is larger than the secret capacity C_S. In this paper, however, we showed that if we use multiplex wiretap channel coding to transmit plural independent messages at the same time, every message can be sent securely if the total coding rate is less than the channel capacity C. In addition to the average criteria, we also proved that the coding theorems for multiplex wiretap channel coding hold with the maximum criteria for the error probability and the security measure of the variational distance.

Finally, it is worth noting that the idea of secure multiplex coding can be applied to several information-theoretic crypto-systems. Actually, strongly secure ramp secret sharing schemes [15], strongly secure linear network coding [16], and multiplex coding for bit commitment [17] are constructed based on ideas similar to the secure multiplex coding treated in this paper.

Acknowledgment

The authors would like to thank the reviewers and the associate editor for their valuable comments.

References

[1] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, John Wiley & Sons, 2006.
[3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 339–348, 1978.
[4] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag, 2003.
[6] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, 1993.
[7] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 5–29, 1989.
[8] I. Devetak, "The private classical information capacity and quantum information capacity of a quantum channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 44–55, 2005.
[9] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1562–1575, 2006.
[10] T. S. Han and S. Verdú, "Spectrum invariancy under output approximation for full-rank discrete memoryless channels," *Problemy Peredachi Informatsii*, vol. 29, no. 2, pp. 9–27, 1993 (*Problem of Information Transmission*, vol. 29, no. 2, pp. 101–118, 1993).
[11] L. H. Ozarow and A. D.Wyner, "Wire-tap channel II," *Advances in Cryptology-Eurocrypt'84*, LNCS 209, pp. 33-50, 1985
[12] A. Barg and G. D.Forney, "Random codes: minimum distance and error exponent," *IEEE Trans. Inform. Theory*, vol. 48, no.9, pp.2568–2573, 2002
[13] Y. Chen and A. J. H. Vinck, "Secrecy coding for the binary symmetric wiretap channel," *Security and Communication Networks*, vol. 4, no. 8, pp.966–978, Aug. 2011
[14] U. Maurer and Stefan Wolf, "Information-theoretic key agreement: from week to strong secrecy for free," *Advances in Cryptology-Eurocrypt 2000, Lecture Notes in Computer Science*, 1807, pp. 351–368, 2000
[15] H. Yamamoto, "Secret sharing system using (k,L,n) threshold scheme," *Trans. of the IECE of Japan*, vol. J68-A, no. 9, pp. 945-952, 1985 (in Japanese), [English translation: *Electronics and Communications in Japan*, Part I, vol. 69, no. 9, pp 46–54, (Scripta Technica, Inc.), 1986]
[16] K. Harada and H. Yamamoto, "Strongly Secure Linear Network Coding," *IEICE Trans. on Fundamentals*, vol. E91, no. 10, pp. 2720-2728, 2008
[17] H. Yamamoto and D. Isami, "Multiplex coding of bit commitment based on a discrete memoryless channel," *Proc. of IEEE-ISIT2007*, pp. 721–725, June 24–29, Nice, France, 2007.

Daisuke Kobayashi was born in Chiba, Japan, in 1979. He received the B. Eng. and M. Eng. degrees in 2003 and 2005, respectively, from the University of Tokyo. He has been with NTT DATA Corporation since 2005.

Hirosuke Yamamoto (S'77-M'80-SM'03-F'11) was born in Wakayama, Japan, in 1952. He received the B.E. degree from Shizuoka University, Shizuoka, Japan, in 1975 and the M.E. and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 1977 and 1980, respectively, all in electrical engineering. In 1980, he joined Tokushima University. He was an Associate Professor at Tokushima University from 1983 to 1987, the University of Electro-Communications from 1987 to 1993, and the University of Tokyo from 1993 to 1999. Since 1999, he has been a Professor at the University of Tokyo and is currently with the Department of Complexity Science and Engineering at the university. In 1989-1990, he was a Visiting Scholar at the Information Systems Laboratory, Stanford University, Stanford, CA. His research interests are in Shannon theory, data compression algorithms, and cryptology.

Dr. Yamamoto served as the Chair of IEEE Information Theory Society Japan Chapter in 2002-2003, the TPC Co-Chair of the ISITA2004, the TPC Chair of the ISITA2008, the president of the SITA (Society of Information Theory and its Applications) in 2008-2009, the president of the ESS (Engineering Sciences Society) of IEICE in 2012-2013, an Associate Editor for Shannon Theory, the IEEE TRANSACTIONS ON INFORMATION THEORY in 2007-2010, Editor-in-Chief for the IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences in 2009-2011. He is a Fellow of the IEICE.

Tomohiro Ogawa was born in Kanagawa, Japan, in 1969. He received the B. Eng. and M. Eng. degrees in 1995 and 1997, respectively, from the University of Tokyo and the Dr. Eng. degree from the University of Electro-Communications in 2000.

He worked at the University of Tokyo from 2000 to 2005, at Japan Science and Technology Agency from 2005 to 2008, and since then he has been with the University of Electro-Communications. His research interests include quantum information theory and information geometry.