Entropy **2013**, *15*, 445-473; doi:10.3390/e15020445

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Wiretap Channel with Action-Dependent Channel State Information

Bin Dai [1,2,*], A. J. Han Vinck [3], Yuan Luo [2] and Xiaohu Tang [1]

[1] School of Information Science and Technology, Southwest JiaoTong University, Northbound Section Second Ring Road 111, Chengdu, China; E-Mail: xhutang@home.swjtu.edu.cn

[2] Computer Science and Engineering Department, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai, China; E-Mail: yuanluo@sjtu.edu.cn

[3] Institute for Experimental Mathematics, Duisburg-Essen University, Ellernstr.29, Essen, Germany; E-Mail: vinck@iem.uni-due.de

* Author to whom correspondence should be addressed; E-Mail: daibinsjtu@gmail.com; Tel.: +86-21-3420-5477.

Received: 23 November 2012; in revised form: 9 January 2013 / Accepted: 17 January 2013 / Published: 28 January 2013

Abstract: In this paper, we investigate the model of wiretap channel with action-dependent channel state information. Given the message to be communicated, the transmitter chooses an action sequence that affects the formation of the channel states, and then generates the channel input sequence based on the state sequence and the message. The main channel and the wiretap channel are two discrete memoryless channels (DMCs), and they are connected with the legitimate receiver and the wiretapper, respectively. Moreover, the transition probability distribution of the main channel depends on the channel state. Measuring wiretapper's uncertainty about the message by equivocation, inner and outer bounds on the capacity-equivocation region are provided both for the case where the channel inputs are allowed to depend non-causally on the state sequence and the case where they are restricted to causal dependence. Furthermore, the secrecy capacities for both cases are bounded, which provide the best transmission rate with perfect secrecy. The result is further explained via a binary example.

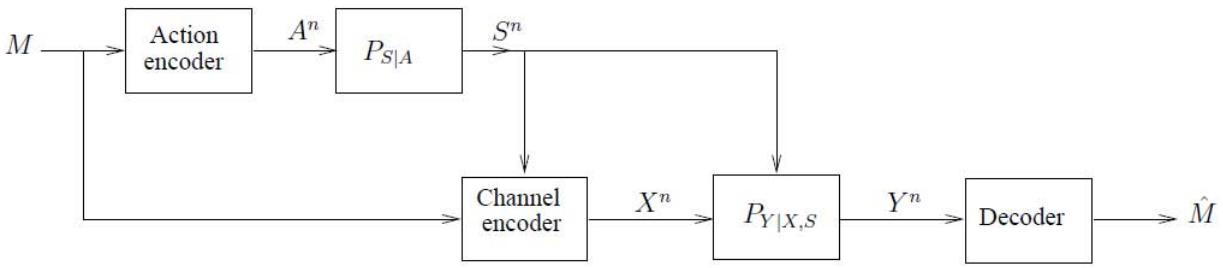
Keywords: action-dependent channel state information; capacity-equivocation region; causal; noncausal; wiretap channel

1. Introduction

Communication through state-dependent channels, with states known at the transmitter, was first investigated by Shannon [1] in 1958. In [1], the capacity of the discrete memoryless channel with causal (past and current) channel state information at the encoder was totally determined. After that, in order to solve the problem of coding for a computer memory with defective cells, Kuznetsov and Tsybakov [2] considered a channel in the presence of non-causal channel state information at the transmitter. They provided some coding techniques without determination of the capacity. The capacity was found in 1980 by Gel'fand and Pinsker [3]. Furthermore, Costa [4] investigated a power constrained additive noise channel, where part of the noise is known at the transmitter as side information. This channel is also called dirty paper channel. The assumption in these seminar papers, as well as in the work on communication with state dependent channels that followed, is that the channel states are generated by nature, and can not be affected or controlled by the communication system.

In 2009, Weissman [5] revisited the above problem setting for the case where the transmitter can take actions that affect the formation of the states, see Figure 1. Specifically, Weissman considered a communication system where encoding is in two parts: given the message, an action sequence is created. The actions affect the formation of the channel states, which are accessible to the transmitter when producing the channel input sequence. The capacity of this model is totally determined both for the case where the channel inputs are allowed to depend non-causally on the state sequence and the case where they are restricted to causal dependence. This framework captures various new channel coding scenarios that may arise naturally in recording for magnetic storage devices or coding for computer memories with defects.

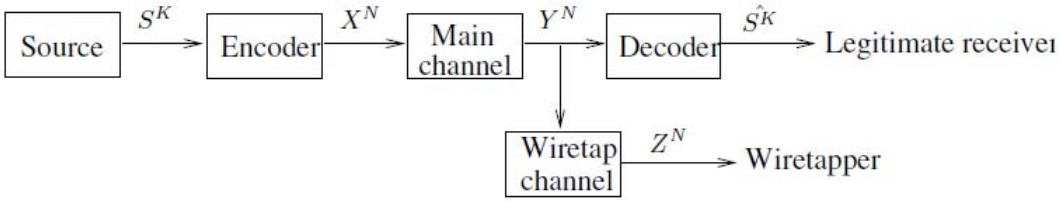
Figure 1. Channel with action-dependent states.



Transmission of confidential messages has been studied in the literature of several classes of channels. Wyner, in his well-known paper on the wiretap channel [6], studied the problem how to transmit the confidential messages to the legitimate receiver via a degraded broadcast channel, while keeping the wiretapper as ignorant of the messages as possible, see Figure 2. Measuring the uncertainty of the wiretapper by equivocation, the capacity-equivocation region was established. Furthermore, the secrecy capacity was also established, which provided the maximum transmission rate with perfect secrecy. After the publication of Wyner's work, Csiszár and Körner [7] investigated a more general situation: the broadcast channels with confidential messages (BCC). In this model, a common message and a confidential message were sent through a general broadcast channel. The common message was assumed to be decoded correctly by the legitimate receiver and the wiretapper, while the confidential message was only allowed to be obtained by the legitimate receiver. This model is also a generalization of [8], where

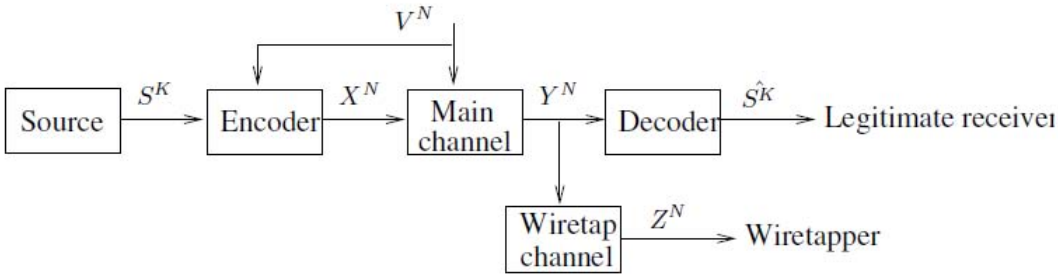
no confidentiality condition is imposed. The capacity-equivocation region and the secrecy capacity region of BCC [7] were totally determined, and the results were also a generalization of those in [6]. Based on Wyner's work, Leung-Yan-Cheong and Hellman studied the Gaussian wiretap channel(GWC) [9], and showed that its secrecy capacity was the difference between the main channel capacity and the overall wiretap channel capacity (the cascade of main channel and wiretap channel).

Figure 2. Wiretap channel.



Inspired by the above works, Mitrpant *et al.* [10] studied transmission of confidential messages in the channels with channel state information (CSI). In [10], an inner bound on the capacity-equivocation region was provided for the Gaussian wiretap channel with CSI. Furthermore, Chen *et al.* [11] investigated the discrete memoryless wiretap channel with noncausal CSI (see Figure 3), and also provided an inner bound on the capacity-equivocation region. Note that the coding scheme of [11] is a combination of those in [3,6] Based on the work of [11], Dai [12] provided an outer bound on the wiretap channel with noncausal CSI, and determined the capacity-equivocation region for the model of wiretap channel with memoryless CSI, where the memoryless means that at the i-th time, the output of the channel encoder depends only on the i-th time CSI.

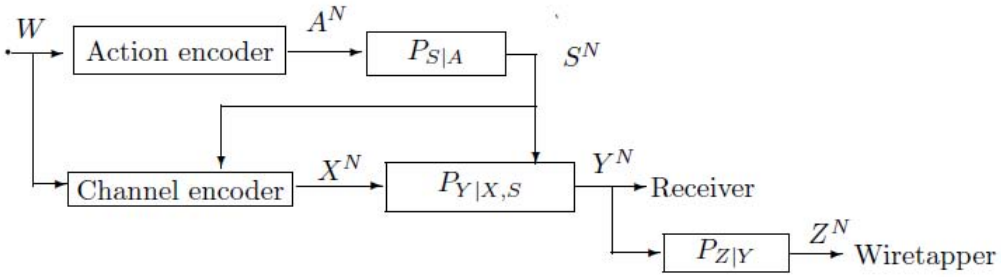
Figure 3. Wiretap channel with noncausal channel state information.



In this paper, we study the wiretap channel with action-dependent channel state information, see Figure 4. Concretely, the transmitted message W is firstly encoded as an action sequence A^N, and A^N is the input of a discrete memoryless channel (DMC). The output of this DMC is the channel state sequence S^N. Then, the transmitted message W and the state sequence S^N are encoded as X^N. The main channel is a DMC with inputs X^N and S^N, and output Y^N. The wiretap channel is also a DMC with input Y^N and output Z^N. Since the action-dependent state captures various new coding scenarios for channels with a rewrite option that may arise naturally in storage for computer memories with defects or in magnetic recoding, it is natural to ask: how about the security of these channel models in the presence of a wiretapper? Measuring wiretapper's uncertainty about the transmitted message by equivocation, the inner and outer bounds on the capacity-equivocation region of the model of Figure 4 are provided both

for the case where the channel input is allowed to depend non-causally on the state sequence and the case where it is restricted to causal dependence.

Figure 4. Wiretap channel with action-dependent channel state information.



In this paper, random variables, sample values and alphabets are denoted by capital letters, lower case letters and calligraphic letters, respectively. A similar convention is applied to the random vectors and their sample values. For example, U^N denotes a random N-vector $(U_1, ..., U_N)$, and $u^N = (u_1, ..., u_N)$ is a specific vector value in $\mathcal{U}^N$ that is the Nth Cartesian power of $\mathcal{U}$. U_i^N denotes a random $N-i+1$-vector $(U_i, ..., U_N)$, and $u_i^N = (u_i, ..., u_N)$ is a specific vector value in $\mathcal{U}_i^N$. Let $p_V(v)$ denote the probability mass function $Pr\{V = v\}$. Throughout the paper, the logarithmic function is to the base 2.

The remainder of this paper is organized as follows. In Section 2, we present the basic definitions and the main result on the capacity-equivocation region of wiretap channel with action-dependent channel state information. In Section 3, we provide a binary example of the model of Figure 4. Final conclusions are presented in Section 4.

2. Notations, Definitions and the Main Results

In this section, the model of Figure 4 is considered into two parts. The model of Figure 4 with noncausal channel state information is described in Subsection 2.1, and the causal case is described in Subsection 2.2, see the followings.

2.1. The Model of Figure 4 with Noncausal Channel State Information

In this subsection, a description of the wiretap channel with noncausal action-dependent channel state information is given by Definition 1 to Definition 6. The inner and outer bounds on the capacity-equivocation region $\mathcal{C}^n$ composed of all achievable (R, R_e) pairs are given in Theorem 1 and Theorem 2, respectively, where the achievable (R, R_e) pair is defined in Definition 6.

Definition 1 *(**Action encoder**) The message W take values in $\mathcal{W}$, and it is uniformly distributed over its range. The action encoder is a deterministic mapping:*

$$f_1^N : \mathcal{W} \to \mathcal{A}^N \tag{1}$$

The input of the action encoder is W, while the output is A^N.

The channel state sequence S^N is generated by a DMC with input A^N and output S^N. The transition probability distribution is given by

$$p_{S^N|A^N}(s^N|a^N) = \prod_{i=1}^{N} p_{S_i|A_i}(s_i|a_i) \tag{2}$$

Note that the components of the state sequence S^N may not be i.i.d. random variables, and this is due to the fact that A^N is not i.i.d. generated.

The transmission rate of the message is $\frac{\log\|\mathcal{W}\|}{N}$.

Definition 2 *(Channel encoder) The inputs of the channel encoder are W and S^N, while the output is X^N. The channel encoder f_2^N is a matrix of conditional probabilities $f_2^N(x^N|w, s^N)$, where $x^N \in \mathcal{X}^N$, $w \in \mathcal{W}$, $s^N \in \mathcal{S}^N$, $\sum_{x^N} f_2^N(x^N|w, s^N) = 1$, and $f_2^N(x^N|w, s^N)$ is the probability that the message w and the channel state sequence s^N are encoded as the channel input x^N.*

Since the channel encoder knows the state sequence s^N in a noncausal manner, at the i-th time ($1 \leq i \leq N$), the channel encoder $f_{2,i}^N$ is a matrix of conditional probabilities $f_{2,i}^N(x_i|w, s^N)$, where $x_i \in \mathcal{X}$, $w \in \mathcal{W}$, $s^N \in \mathcal{S}^N$, $\sum_{x_i} f_{2,i}^N(x_i|w, s^N) = 1$, and $f_{2,i}^N(x_i|w, s^N)$ is the probability that the message w and the channel state sequence s^N are encoded as the i-th time channel input x_i.

The transmission rate of the message is $\frac{\log\|\mathcal{W}\|}{N}$.

Definition 3 *(Main channel) The main channel is a DMC with finite input alphabet $\mathcal{X} \times \mathcal{S}$, finite output alphabet $\mathcal{Y}$, and transition probability $Q_M(y|x, s)$, where $x \in \mathcal{X}, s \in \mathcal{S}, y \in \mathcal{Y}$. $Q_M(y^N|x^N, s^N) = \prod_{n=1}^{N} Q_M(y_n|x_n, s_n)$. The inputs of the main channel are X^N and S^N, while the output is Y^N.*

Definition 4 *(Wiretap channel) The wiretap channel is also a DMC with finite input alphabet $\mathcal{Y}$, finite output alphabet $\mathcal{Z}$, and transition probability $Q_W(z|y)$, where $y \in \mathcal{Y}, z \in \mathcal{Z}$. The input and output of the wiretap channel are Y^N and Z^N, respectively. The equivocation to the wiretapper is defined as*

$$\Delta = \frac{H(W|Z^N)}{N} \tag{3}$$

The cascade of the main channel and the wiretap channel is another DMC with transition probability

$$Q_{MW}(z|x, s) = \sum_{y \in \mathcal{Y}} Q_W(z|y) Q_M(y|x, s) \tag{4}$$

Note that, $(X^N, S^N) \to Y^N \to Z^N$ and $W \to A^N \to S^N$ are two Markov chains in the model of Figure 4.

Definition 5 *(Decoder) The decoder for the legitimate receiver is a mapping $f_{D1} : \mathcal{Y}^N \to \mathcal{W}$, with input Y^N and output $\widehat{W}$. Let P_e be the error probability of the receiver , and it is defined as $Pr\{W \neq \widehat{W}\}$.*

Definition 6 *(Achievable (R, R_e) pair in the model of Figure 4) A pair (R, R_e) (where $R, R_e > 0$) is called achievable if, for any $\epsilon > 0$ (where ϵ is an arbitrary small positive real number and $\epsilon \to 0$), there exists channel encoders-decoders (N, P_e) such that*

$$\lim_{N\to\infty} \frac{\log \parallel \mathcal{W} \parallel}{N} = R, \lim_{N\to\infty} \Delta \geq R_e P_e \leq \epsilon \tag{5}$$

The capacity-equivocation region $\mathcal{R}^n$ is a set composed of all achievable (R, R_e) pairs. Inner and outer bounds on $\mathcal{R}^n$ are respectively provided in the following Theorem 1 and Theorem 2. Theorem 1 and Theorem 2 are respectively proved in Section A and Section B.

Theorem 1 ***(Inner bound)*** *A single-letter characterization of the region $\mathcal{R}^{ni}$ is as follows,*

$$\begin{aligned}\mathcal{R}^{(ni)} = \{(R, R_e) : 0 \leq R_e \leq R \\ R \leq I(U;Y) - I(U;S|A) \\ R_e \leq I(U;Y) - I(U;Z) \\ R_e \leq H(A|Z)\}\end{aligned}$$

where $p_{UASXYZ}(u,a,s,x,y,z) = p_{Z|Y}(z|y)p_{Y|X,S}(y|x,s)p_{UAXS}(u,a,x,s)$, *which implies that* $(A,U) \to (X,S) \to Y \to Z$.

The region $\mathcal{R}^{(ni)}$ satisfies $\mathcal{R}^{(ni)} \subseteq \mathcal{R}^{(n)}$.

Remark 1 *There are some notes on Theorem 1, see the following.*

- *The formula $R_e \leq H(A|Z)$ in Theorem 1 implies that the wiretapper obtains the information about the message not only from the codeword transmitted in the channels, but also from the action sequence a^N. If the wiretapper knows a^N, he knows the corresponding message.*
- *The region $\mathcal{R}^{(ni)}$ is convex, and the proof is directly obtained by introducing a time sharing random variable into Theorem 1, and therefore, we omit the proof here.*
- *The range of the random variable U satisfies*

  $$\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{A}\|\|\mathcal{S}\| + 2$$

  The proof is in Section C.
- *Without the equivocation parameter, the capacity of the main channel is given by*

  $$C_M = \max_{p_{X|U,S}(x|u,s)p_{U|A,S}(u|a,s)p_A(a)} (I(U;Y) - I(U;S|A)) \tag{6}$$

  The formula (6) is proved by Weissman [5], and it is omitted here.
- Secrecy capacity

  The points in $\mathcal{R}^{(ni)}$ for which $R_e = R$ are of considerable interest, which imply the perfect secrecy $H(W) = H(W|Z^N)$. Clearly, we can easily bound the secrecy capacity C_s^n of the model of Figure 4 with noncausal channel state information by

  $$C_s^n \geq \max_{p_{UAXS}(u,a,x,s)} \min\{I(U;Y) - I(U;Z), I(U;Y) - I(U;S|A), H(A|Z)\} \tag{7}$$

Proof 1 (Proof of (7)) *Substituting $R_e = R$ into the region $\mathcal{R}^{(ni)}$ in Theorem 1, we have*

$$R \leq I(U;Y) - I(U;Z) \tag{8}$$
$$R \leq I(U;Y) - I(U;S|A) \tag{9}$$
$$R \leq H(A|Z) \tag{10}$$

Note that the pair $(R = \max\min\{I(U;Y) - I(U;Z), I(U;Y) - I(U;S|A), H(A|Z)\}, R_e = R)$ *is achievable, and therefore, the secrecy capacity* $C_s^{(n)} \geq \max\min\{I(U;Y) - I(U;Z), I(U;Y) - I(U;S|A), H(A|Z)\}$. *Thus the proof is completed.*

Theorem 2 ***(Outer bound)*** *A single-letter characterization of the region* $\mathcal{R}^{no}$ *is as follows,*

$$\begin{aligned}\mathcal{R}^{(no)} = \{(R, R_e) : 0 \leq R_e \leq R \\ R \leq I(U;Y) - I(U;S|A) \\ R_e \leq I(U;Y) - I(K;Z|V)\}\end{aligned}$$

where $p_{UKVASXYZ}(u,k,v,a,s,x,y,z) = p_{Z|Y}(z|y)p_{Y|X,S}(y|x,s)p_{X|U,S}(x|u,s)p_{V|K}(v|k)p_{K|U}(k|u)p_{U,A,S}(u,a,s)$, *which implies that* $(A,U,K,V) \to (X,S) \to Y \to Z$ *and* $V \to K \to U \to Y \to Z$ *are two Markov chains.*

The region $\mathcal{R}^{(no)}$ *satisfies* $\mathcal{R}^{(n)} \subseteq \mathcal{R}^{(no)}$.

Remark 2 *There are some notes on Theorem 2, see the following.*

- *The region* $\mathcal{R}^{(no)}$ *is convex, and the proof is similar to that of Theorem 1. Therefore, we omit the proof here.*
- *The ranges of the random variables* U, V *and* K *satisfy*

$$\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{A}\|\|\mathcal{S}\| + 1$$

$$\|\mathcal{V}\| \leq \|\mathcal{X}\|\|\mathcal{A}\|\|\mathcal{S}\|$$

$$\|\mathcal{K}\| \leq \|\mathcal{X}\|^2\|\mathcal{A}\|^2\|\mathcal{S}\|^2$$

  The proof is in Section D.
- *Observing the formula* $R_e \leq I(U;Y) - I(K;Z|V)$ *in Theorem 2, we have*

$$\begin{aligned} I(U;Y) - I(K;Z|V) &=^{(a)} I(U;Y) - H(Z|V) + H(Z|K) \\ &\geq I(U;Y) - H(Z) + H(Z|K) \\ &\geq I(U;Y) - H(Z) + H(Z|K,U) \\ &=^{(b)} I(U;Y) - H(Z) + H(Z|U) = I(U;Y) - I(U;Z) \end{aligned} \quad (11)$$

  where (a) is from the fact that $V \to K \to Z$, *and (b) is from the Markov chain* $K \to U \to Y \to Z$. *Then it is easy to see that* $\mathcal{R}^{(ni)} \subseteq \mathcal{R}^{(no)}$.
- *The secrecy capacity* C_s^n *of the model of Figure 4 with noncausal channel state information is upper bounded by*

$$C_s^n \leq \max_{p_{X,U,K,V,A,S}(x,u,k,v,a,s)} \min\{I(U;Y) - I(K;Z|V), I(U;Y) - I(U;S|A)\} \quad (12)$$

  The upper bound is easily obtained by substituting $R_e = R$ *into the region* $\mathcal{R}^{(no)}$ *in Theorem 2, and therefore, we omit the proof here.*

2.2. The Model of Figure 4 with Causal Channel State Information

The model of Figure 4 with causal channel state information is similar to the model with noncausal channel state information in Subsection 2.1, except that the state sequence S^N in Definition 1 is known to the channel encoder in a causal manner, *i.e.*, at the i-th time ($1 \leq i \leq N$), the output of the encoder $x_i = f_{2,i}(w, s^i)$, where $s^i = (s_1, s_2, ..., s_i)$ and $f_{2,i}$ is the probability that the message w and the state s^i are encoded as the channel input x_i at time i. Define

$$f^N(x^N|w, s^N) = \prod_{i=1}^{N} f_i(x_i|w, s^i) \tag{13}$$

Inner and outer bounds on the capacity-equivocation region $\mathcal{R}^c$ for the model of Figure 4 with causal channel state information are respectively provided in the following Theorem 3 and Theorem 4, and they are proved in Section E and Section F.

Theorem 3 ***(Inner bound)*** *A single-letter characterization of the region $\mathcal{R}^{ci}$ is as follows,*

$$\begin{aligned}\mathcal{R}^{(ci)} = \{&(R, R_e) : 0 \leq R_e \leq R\\ &R \leq I(U;Y)\\ &R_e \leq I(U;Y) - I(U;Z)\\ &R_e \leq H(A|Z)\}\end{aligned}$$

where $p_{UASXYZ}(u, a, s, x, y, z) = p_{Z|Y}(z|y)p_{Y|X,S}(y|x, s)p_{X|U,S}(x|u, s)p_{S|A}(s|a)p_{UA}(u, a)$***, which implies that*** $(A, U) \to (X, S) \to Y \to Z$ ***and*** $U \to A \to S$***.***

The region $\mathcal{R}^{(ci)}$ satisfies $\mathcal{R}^{(ci)} \subseteq \mathcal{R}^{(c)}$.

Remark 3 *There are some notes on Theorem 3, see the following.*

- *The region $\mathcal{R}^{(ci)}$ is convex.*
- *The range of the random variable U satisfies*

$$\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{A}\|\|\mathcal{S}\| + 1$$

  The proof is similar to that in Theorem 1, and it is omitted here.
- *Without the equivocation parameter, the capacity of the main channel is given by*

$$C_M^* = \max_{p_{X|U,S}(x|u,s)p_{U,A}(u,a)} I(U;Y) \tag{14}$$

  The formula (14) is proved by Weissman [5], and it is omitted here.
- Secrecy capacity

  The points in $\mathcal{R}^{(ci)}$ for which $R_e = R$ are of considerable interest, which imply the perfect secrecy $H(W) = H(W|Z^N)$. Clearly, we can easily bound the secrecy capacity $C_s^{(c)}$ of the model of Figure 4 with causal channel state information by

$$C_s^{(c)} \geq \max_{p_{X|U,S}(x|u,s)p_{UA}(u,a)} \min\{I(U;Y) - I(U;Z), H(A|Z)\} \tag{15}$$

Proof 2 (Proof of (15)) *Substituting $R_e = R$ into the region $\mathcal{R}^{(ci)}$ in Theorem 3, we have*

$$R \leq I(U;Y) - I(U;Z) \tag{16}$$
$$R \leq I(U;Y) \tag{17}$$
$$R \leq H(A|Z) \tag{18}$$

Note that the pair $(R = \max\min\{I(U;Y) - I(U;Z), H(A|Z)\}, R_e = R)$ is achievable, and therefore, the secrecy capacity $C_s^{(c)} \geq \max\min\{I(U;Y) - I(U;Z), H(A|Z)\}$. Thus the proof is completed.

Theorem 4 ***(Outer bound)*** *A single-letter characterization of the region $\mathcal{R}^{co}$ is as follows,*

$$\begin{aligned}\mathcal{R}^{(co)} = \{(R, R_e) : 0 \leq R_e \leq R \\ R \leq I(U;Y) \\ R_e \leq I(U;Y) - I(K;Z|V)\}\end{aligned}$$

where $p_{UKVASXYZ}(u,k,v,a,s,x,y,z) = p_{Z|Y}(z|y)p_{Y|X,S}(y|x,s)p_{X|U,S}(x|u,s)p_{V|K}(v|k)p_{K|U}(k|u)$ $p_{U,A,S}(u,a,s)$, which implies that $(A,U,K,V) \to (X,S) \to Y \to Z$ and $V \to K \to U \to Y \to Z$ are two Markov chains.

The region $\mathcal{R}^{(co)}$ satisfies $\mathcal{R}^{(c)} \subseteq \mathcal{R}^{(co)}$.

Remark 4 *There are some notes on Theorem 4, see the following.*

- *The region $\mathcal{R}^{(co)}$ is convex.*
- *The ranges of the random variables U, V and K satisfy*

$$\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{A}\|\|\mathcal{S}\|$$
$$\|\mathcal{V}\| \leq \|\mathcal{X}\|\|\mathcal{A}\|\|\mathcal{S}\|$$
$$\|\mathcal{K}\| \leq \|\mathcal{X}\|^2\|\mathcal{A}\|^2\|\mathcal{S}\|^2$$

  The proof is similar to that in Section D, and it is omitted here.
- *The secrecy capacity $C_s^{(c)}$ of the model of Figure 4 with causal channel state information is upper bounded by*

$$C_s^{(c)} \leq \max_{p_{X|U,S}(x|u,s)p_{U,K,V,A,S}(u,k,v,a,s)} I(U;Y) - I(K;Z|V) \tag{19}$$

  The upper bound is easily obtained by substituting $R_e = R$ into the region $\mathcal{R}^{(co)}$ in Theorem 4, and therefore, we omit the proof here.

3. A Binary Example for the Model of Figure 4 with Causal Channel State Information

In this section, we calculate the bound on secrecy capacity of a special case of the model of Figure 4 with causal channel state information.

Suppose that the channel state information S^N is available at the channel encoder in a casual manner. Meanwhile, the random variables X, Y and Z take values in $\{0,1\}$, and the transition probability of the main channel is defined as follows:

When $s=0$,

$$p_{Y|X,S}(y|x,s=0)=\begin{cases}1-p, & \text{if } y=x,\\ p, & \text{otherwise.}\end{cases} \tag{20}$$

When $s=1$,

$$p_{Y|X,S}(y|x,s=1)=\begin{cases}p, & \text{if } y=x,\\ 1-p, & \text{otherwise.}\end{cases} \tag{21}$$

The wiretap channel is a BSC (binary symmetric channel) with crossover probability q, *i.e.*,

$$p_{Z|Y}(z|y)=\begin{cases}1-q, & \text{if } y=x,\\ q, & \text{otherwise.}\end{cases} \tag{22}$$

The channel for generating the state sequence S^N is a BSC with crossover probability r, *i.e.*,

$$p_{S|A}(s|a)=\begin{cases}1-r, & \text{if } y=x,\\ r, & \text{otherwise.}\end{cases} \tag{23}$$

From Remark 3 and Remark 4 we know that the secrecy capacity for the causal case is bounded by

$$\max\min\{I(U;Y)-I(U;Z),H(A|Z)\}\le C_s^c\le\max(I(U;Y)-I(K;Z|V))\le^{(a)}\max I(U;Y). \tag{24}$$

Note that in (a), "=" is achieved if $V=K$. Moreover, $\max I(U;Y)$, $\max H(A|Z)$ and $\max(I(U;Y)-I(U;Z))$ are achieved if A is a function of U and X is a function of U and S, and this is similar to the argument in [5]. Define $a=g(u)$ and $x=f(u,s)$, then (24) can be written as

$$\max_{f,g,p_U(u)}\min\{I(U;Y)-I(U;Z),H(A|Z)\}\le C_s^c\le\max_{f,g,p_U(u)}I(U;Y) \tag{25}$$

and this is because the joint probability distribution $p_{AUSXYZ}(a,u,s,x,y,z)$ can be calculated by

$$p_{AUSXYZ}(a,u,s,x,y,z)=p_{Z|Y}(z|y)p_{Y|X,S}(y|x,s)1_{x=f(u,s)}p_{S|A}(s|a)1_{a=g(u)}p_U(u) \tag{26}$$

Now it remains to calculate the characters $\max_{f,g,p_U(u)}(I(U;Y)-I(U;Z))$, $\max_{f,g,p_U(u)}H(A|Z)$ and $\max_{f,g,p_U(u)}I(U;Y)$, see the remaining of this section.

Let U take values in $\{0,1\}$. The probability of U is defined as follows. $p_U(0)=\alpha$ and $p_U(1)=1-\alpha$. In addition, there are 16 kinds of f and 4 kinds of g. Define

$$f^{(1)}(u,s):\begin{cases}00\to0,01\to0,\\ 10\to0,11\to0.\end{cases}\quad f^{(2)}(u,s):\begin{cases}00\to0,01\to0,\\ 10\to0,11\to1.\end{cases} \tag{27}$$

$$f^{(3)}(u,s):\begin{cases}00\to0,01\to0,\\ 10\to1,11\to0.\end{cases}\quad f^{(4)}(u,s):\begin{cases}00\to0,01\to0,\\ 10\to1,11\to1.\end{cases} \tag{28}$$

$$f^{(5)}(u,s):\begin{cases}00\to0,01\to1,\\ 10\to0,11\to0.\end{cases}\quad f^{(6)}(u,s):\begin{cases}00\to0,01\to1,\\ 10\to0,11\to1.\end{cases} \tag{29}$$

$$f^{(7)}(u,s):\begin{cases}00\to0,01\to1,\\ 10\to1,11\to0.\end{cases}\quad f^{(8)}(u,s):\begin{cases}00\to0,01\to1,\\ 10\to1,11\to1.\end{cases} \tag{30}$$

$$f^{(9)}(u,s):\begin{cases}00\to 1,01\to 0,\\ 10\to 0,11\to 0.\end{cases}\quad f^{(10)}(u,s):\begin{cases}00\to 1,01\to 0,\\ 10\to 0,11\to 1.\end{cases} \tag{31}$$

$$f^{(11)}(u,s):\begin{cases}00\to 1,01\to 0,\\ 10\to 1,11\to 0.\end{cases}\quad f^{(12)}(u,s):\begin{cases}00\to 1,01\to 0,\\ 10\to 1,11\to 1.\end{cases} \tag{32}$$

$$f^{(13)}(u,s):\begin{cases}00\to 1,01\to 1,\\ 10\to 0,11\to 0.\end{cases}\quad f^{(14)}(u,s):\begin{cases}00\to 1,01\to 1,\\ 10\to 0,11\to 1.\end{cases} \tag{33}$$

$$f^{(15)}(u,s):\begin{cases}00\to 1,01\to 1,\\ 10\to 1,11\to 0.\end{cases}\quad f^{(16)}(u,s):\begin{cases}00\to 1,01\to 1,\\ 10\to 1,11\to 1.\end{cases} \tag{34}$$

$$g^{(1)}(u):\begin{cases}0\to 0,\\ 1\to 0.\end{cases}\quad g^{(2)}(u):\begin{cases}0\to 0,\\ 1\to 1.\end{cases} \tag{35}$$

$$g^{(3)}(u):\begin{cases}0\to 1,\\ 1\to 0.\end{cases}\quad g^{(4)}(u):\begin{cases}0\to 1,\\ 1\to 1.\end{cases} \tag{36}$$

The character $I(U;Y)$ depends on the joint probability mass functions $p_{UY}(u,y)$, and we have

$$\begin{aligned} p_{UY}(u,y) &= \sum_{x,s,a} p_{UYXSA}(u,y,x,s,a) \\ &= \sum_{x,s,a} p_{Y|XS}(y|x,s)p_{X|U,S}(x|u,s)p_U(u)p_{A|U}(a|u)p_{S|A}(s|a) \end{aligned} \tag{37}$$

The character $I(U;Z)$ depends on the joint probability mass functions $p_{UZ}(u,z)$, and we have

$$\begin{aligned} p_{UZ}(u,z) &= \sum_{y} p_{UYZ}(u,y,z) \\ &= \sum_{y} p_{Z|Y}(z|y)p_{U,Y}(u,y) \end{aligned} \tag{38}$$

By choosing the above f, g and α, we find that

$$\max_{f,g,p_U(u)}(I(U;Y)-I(U;Z)) = \max\{h(p\star q)-h(p), \frac{h(q\star(r\star p))-h(p\star r)}{2r}-(\frac{1}{2r}-1)(h(p\star q)-h(p))\} \tag{39}$$

where $p\star q = p+q-2pq$. Moreover, $h(p\star q)-h(p)$ is achieved when $f=f^{(7)}$, $g=g^{(2)}$ and $\alpha=\frac{1}{2}$, and $\frac{h(q\star(r\star p))-h(p\star r)}{2r}-(\frac{1}{2r}-1)(h(p\star q)-h(p))$ is achieved when $f=f^{(2)}$, $g=g^{(2)}$ and $\alpha=\frac{1}{2}$.

Moreover,

$$\max_{f,g,p_U(u)} H(A|Z) = h(p\star q) \tag{40}$$

where $p\star q = p+q-2pq$. Moreover, $h(p\star q)$ is achieved when $f=f^{(7)}$, $g=g^{(2)}$ and $\alpha=\frac{1}{2}$.

In addition,

$$\max_{f,g,p_U(u)} I(U;Y) = 1-h(p) \tag{41}$$

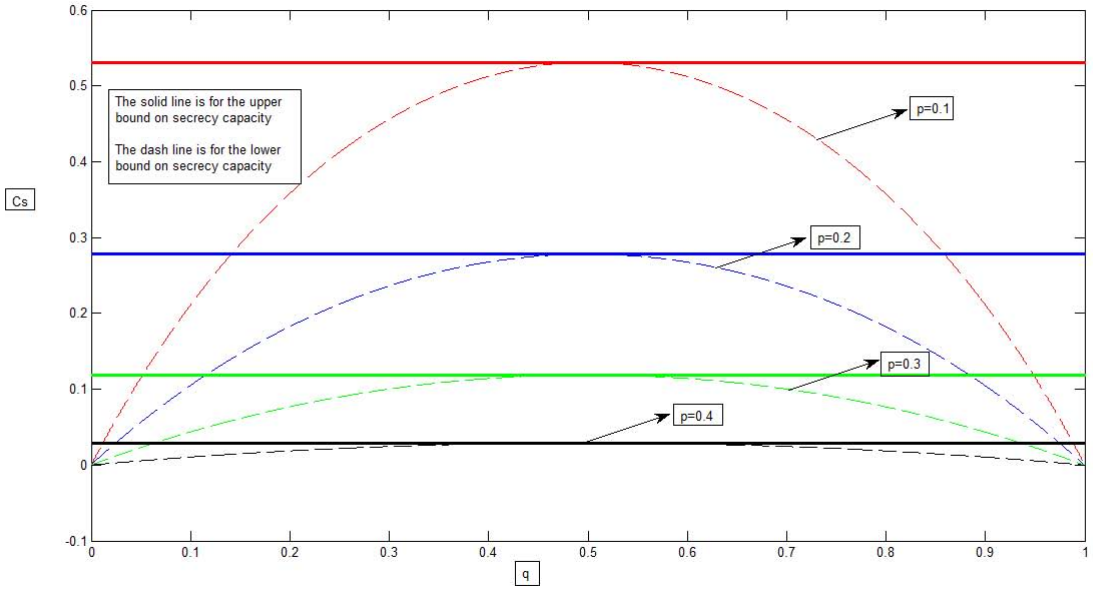
and "=" is achieved if $f=f^{(7)}$, $g=g^{(2)}$ and $\alpha=\frac{1}{2}$.

It is easy to see that $\max_{f,g,p_U(u)} H(A|Z) = h(p \star q) \geq \max_{f,g,p_U(u)}(I(U;Y) - I(U;Z))$ and therefore, the secrecy capacity for the causal case is bounded by

$$\max\{h(p \star q) - h(p), \frac{h(q \star (r \star p)) - h(p \star r)}{2r} - (\frac{1}{2r} - 1)(h(p \star q) - h(p))\} \leq C_s^c \leq 1 - h(p) \quad (42)$$

The following Figure 5 gives lower and upper bounds on the secrecy capacity of the model of Figure 4 with causal channel state information. It is easy to see that when $q = 0.5$, the lower bound meets with the upper bound, *i.e.*, the secrecy capacity C_s^c satisfies $C_s^c = 1 - h(p)$. This is because when $q = 0.5$, zero leakage is always satisfied and the problem reduces to the problem of coding for channel with causal states. Moreover, when r is fixed, the bounds on secrecy capacity are getting better while p is decreasing.

Figure 5. When r=0.2, lower and upper bounds on the secrecy capacity of the model of Figure 4 with causal channel state information.



4. Conclusions

In this paper, we study the model of the wiretap channel with action-dependent channel state information. Inner and outer bounds on the capacity-equivocation region are provided both for the case where the channel inputs are allowed to depend non-causally on the state sequence and the case where they are restricted to causal dependence. Furthermore, the secrecy capacities for both cases are bounded, which provide the best transmission rate with perfect secrecy. The result is further explained via a binary‘example.

Acknowledgement

The authors would like to thank N. Cai for his help to improve this paper. This work was supported by a sub-project in National Basic Research Program of China under Grant 2012CB316100 on Broadband Mobile Communications at High Speeds, the National Natural Science Foundation of China under

Grant 61271222, and the Research Fund for the Doctoral Program of Higher Education of China (No. 20100073110016).

References

1. Shannon, C.E. Channels with side information at the transmitter. *IBM J. Res. Dev.* **1958**, *2*, 289–293.
2. Kuznetsov, N.V.; Tsybakov, B.S. Coding in memories with defective cells. *Probl. Peredachi Informatsii* **1974**, *10*, 52–60.
3. Gel'fand, S.I.; Pinsker, M.S. Coding for channel with random parameters. *Problems. Control Inf. Theory* **1980**, *9*, 19–31.
4. Costa, M.H.M. Writing on dirty paper. *IEEE Trans. Inf. Theory* **1983**, *29*, 439–441.
5. Weissman, T. Capacity of channels with action-dependent states. *IEEE Trans. Inf. Theory* **2010**, *56*, 5396–5411.
6. Wyner, A.D. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387.
7. Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348.
8. Körner, J.; Marton, K. General broadcast channels with degraded message sets. *IEEE Trans. Inf. Theory* **1977**, *23*, 60–64.
9. Leung-Yan-Cheong, S.K.; Hellman, M.E. The Gaussian wire-tap channel. *IEEE Trans. Inf. Theory* **1978**, *24*, 451–456.
10. Mitrpant, C.; Han Vinck, A.J.; Luo, Y. An achievable region for the gaussian wiretap channel with side information. *IEEE Trans. Inf. Theory* **2006**, *52*, 2181–2190.
11. Chen, Y.; Han Vinck, A.J. Wiretap channel with side information. *IEEE Trans. Inf. Theory* **2008**, *54*, 395–402.
12. Dai, B.; Luo, Y. Some new results on wiretap channel with side information. *Entropy* **2012**, *14*, 1671–1702.
13. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Academic: London, UK, 1981; pp. 123–124.

A. Proof of Theorem 1

In this section, we will show that any pair $(R, R_e) \in \mathcal{R}^{ni}$ is achievable. Gel'fand-Pinsker's binning and Wyner's random binning technique are used in the construction of the code-books.

Now the remainder of this section is organized as follows. The code construction is in Subsection A.1. The proof of achievability is given in Subsection A.2.

A.1. Code Construction

Since $R_e \leq I(U;Y) - I(U;Z)$, $R_e \leq H(A|Z)$ and $R_e \leq R \leq I(U;Y) - I(U;S|A)$, it is sufficient to show that the pair $(R, R_e = \min\{I(U;Y) - \max(I(U;Z), I(U;S|A))), H(A|Z)\}$ is achievable, and note that this implies that $R \geq R_e = \min\{I(U;Y) - \max(I(U;Z), I(U;S|A))), H(A|Z)\}$.

The construction of the code and the proof of achievability are considered into two cases:

- **(Case 1)** If $H(A|Z) \geq I(U;Y) - \max(I(U;Z), I(U;S|A)))$, double binning technique [11] is used in the construction of the code-book.
- **(Case 2)** If $H(A|Z) \leq I(U;Y) - \max(I(U;Z), I(U;S|A)))$, Gel'fand-Pinsker's binning technique [3] is used in the construction of the code-book.

- **(Code construction for Case 1)**
  Given a pair (R, R_e), choose a joint probability mass function $p_{U,A,S,X,Y,Z}(u,a,s,x,y,z)$ such that

$$0 \leq R_e \leq R$$

$$R \leq I(U;Y) - I(U;S|A)$$

$$R_e = I(U;Y) - \max(I(U;Z), I(U;S|A)))$$

  The message set $\mathcal{W}$ satisfies the following condition:

$$\lim_{N\to\infty} \frac{1}{N} \log \parallel \mathcal{W} \parallel = R = I(U;Y) - I(U;S|A) - \gamma \tag{A1}$$

  where γ is a fixed positive real numbers and

$$0 \leq \gamma \leq^{(a)} \max(I(U;Z), I(U;S|A))) - I(U;S|A) \tag{A2}$$

  Note that (a) is from $R \geq R_e = I(U;Y) - \max(I(U;Z), I(U;S|A)))$ and (A1). Let $\mathcal{W} = \{1, 2, ..., 2^{NR}\}$.
  Code-book generation:

  - **(Construction of A^N)**
    Generate 2^{NR} i.i.d. sequences a^N, according to the probability mass function $p_A(a)$. Index each sequence by $i \in \{1, 2, ..., 2^{NR}\}$. For a given message w ($w \in \mathcal{W}$), choose a corresponding $a^N(w)$ as the output of the action encoder.
  - **(Construction of U^N)**
    For the transmitted action sequence $a^N(w)$, generate $2^{N(I(U;Y)-\epsilon_{2,N})}$ ($\epsilon_{2,N} \to 0$ as $N \to \infty$) i.i.d. sequences u^N, according to the probability mass function $p_{U|A}(u_i|a_i(w))$. Distribute these sequences at random into $2^{NR} = 2^{N(I(U;Y)-I(U;S|A)-\gamma)}$ bins such that each bin contains $2^{N(I(U;S|A)+\gamma-\epsilon_{2,N})}$ sequences. Index each bin by $i \in \{1, 2, ..., 2^{NR}\}$. Then place the $2^{N(I(U;S|A)+\gamma-\epsilon_{2,N})}$ sequences in every bin randomly into $2^{N(\max(I(U;S|A),I(U;Z))-I(U;Z)+\epsilon_{3,N})}$ ($\epsilon_{3,N} \to 0$ as $N \to \infty$) subbins such that every subbin contains $2^{N(I(U;S|A)+\gamma-\epsilon_{2,N}-\max(I(U;S|A),I(U;Z))+I(U;Z)-\epsilon_{3,N})}$ sequences. Let J be the random variable to represent the index of the subbin. Index each subbin by $j \in \{1, 2, ..., 2^{N(\max(I(U;S|A),I(U;Z))-I(U;Z)+\epsilon_{3,N})}\}$, *i.e.*,

$$\log \|\mathcal{J}\| = N(\max(I(U;S|A), I(U;Z)) - I(U;Z) + \epsilon_{3,N}). \tag{A3}$$

    Here note that the number of the sequences in every subbin is upper bounded as follows.

$$\begin{aligned} & I(U;S|A) + \gamma - \epsilon_{2,N} - \max(I(U;S|A), I(U;Z)) + I(U;Z) - \epsilon_{3,N} \\ \leq^{(a)} \; & I(U;Z) - \epsilon_{2,N} - \epsilon_{3,N} \end{aligned} \tag{A4}$$

where (a) is from (A2). This implies that

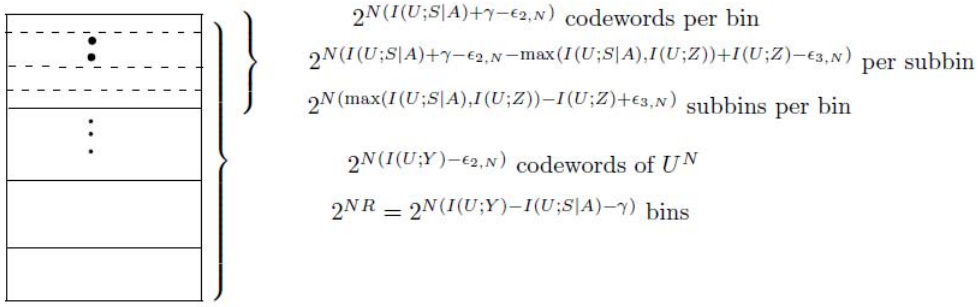
$$\lim_{N\to\infty} H(U^N|W, J, Z^N) = 0. \tag{A5}$$

Note that (A5) can be proved by using Fano's inequality and (A4).

Let s^N be the state sequence generated in response to the action sequence $a^N(w)$. For a given message w ($w \in \mathcal{W}$) and channel state s^N, try to find a sequence $u^N(w, i^*)$ in bin w such that $(u^N(w, i^*), a^N(w), s^N) \in T^N_{UAS}(\epsilon_2)$. If multiple such sequences in bin w exist, choose the one with the smallest index in the bin. If no such sequence exists, declare an encoding error.

Figure A1 shows the construction of U^N for case 1, see the following.

Figure A1. Code-book construction for U^N in Theorem 1 for case 1.



– **(Construction of X^N)** The x^N is generated according to a new discrete memoryless channel (DMC) with inputs u^N, s^N, and output x^N. The transition probability of this new DMC is $p_{X|U,S}(x|u, s)$, which is obtained from the joint probability mass function $p_{U,A,S,X,Y,Z}(u, a, s, x, y, z)$. The probability
$p_{X^N|U^N,S^N}(x^N|u^N, s^N)$ is calculated as follows.

$$p_{X^N|U^N,S^N}(x^N|u^N, s^N) = \prod_{i=1}^{N} p_{X|U,S}(x_i|u_i, s_i) \tag{A6}$$

Decoding:

Given a vector $y^N \in \mathcal{Y}^N$, try to find a sequence $u^N(\hat{w}, \hat{i})$ such that $(u^N(\hat{w}, \hat{i}), a^N(\hat{w}), y^N) \in T^N_{UAY}(\epsilon_3)$. If there exist sequences with the same $\hat{w}$, put out the corresponding $\hat{w}$. Otherwise, *i.e.*, if no such sequence exists or multiple sequences have different message indices, declare a decoding error.

- **(Code construction for Case 2)**

Given a pair (R, R_e), choose a joint probability mass function $p_{U,A,S,X,Y,Z}(u, a, s, x, y, z)$ such that

$$0 \le R_e \le R$$

$$R \le I(U;Y) - I(U;S|A)$$

$$R_e = H(A|Z)$$

The message set $\mathcal{W}$ satisfies the following condition:

$$\lim_{N\to\infty}\frac{1}{N}\log \parallel \mathcal{W} \parallel = R = I(U;Y) - I(U;S|A) - \gamma_1, \tag{A7}$$

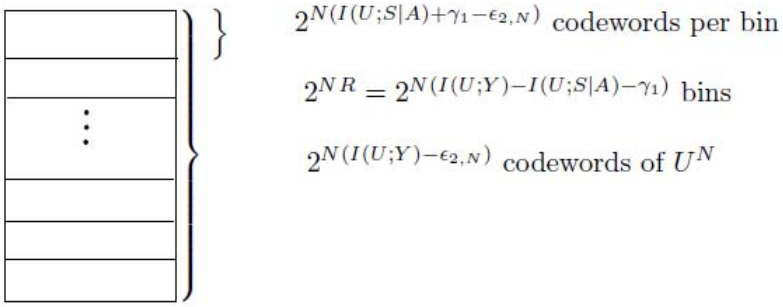
where γ_1 is a fixed positive real numbers and

$$0 \le \gamma_1 \le^{(b)} I(U;Y) - I(U;S|A) - H(A|Z). \tag{A8}$$

Note that (b) is from $R \ge R_e = H(A|Z)$ and (A7). Let $\mathcal{W} = \{1, 2, ..., 2^{NR}\}$.
Code-book generation:

- **(Construction of A^N)**
  Generate 2^{NR} i.i.d. sequences a^N, according to the probability mass function $p_A(a)$. Index each sequence by $i \in \{1, 2, ..., 2^{NR}\}$. For a given message w ($w \in \mathcal{W}$), choose a corresponding $a^N(w)$ as the output of the action encoder.
- **(Construction of U^N)**
  For the transmitted action sequence $a^N(w)$, generate $2^{N(I(U;Y)-\epsilon_{2,N})}$ ($\epsilon_{2,N} \to 0$ as $N \to \infty$) i.i.d. sequences u^N, according to the probability mass function $p_{U|A}(u_i|a_i(w))$. Distribute these sequences at random into $2^{NR} = 2^{N(I(U;Y)-I(U;S|A)-\gamma_1)}$ bins such that each bin contains $2^{N(I(U;S|A)+\gamma_1-\epsilon_{2,N})}$ sequences. Index each bin by $i \in \{1, 2, ..., 2^{NR}\}$.
  Let s^N be the state sequence generated in response to the action sequence $a^N(w)$. For a given message w ($w \in \mathcal{W}$) and channel state s^N, try to find a sequence $u^N(w, i^*)$ in bin w such that $(u^N(w, i^*), a^N(w), s^N) \in T^N_{UAS}(\epsilon_2)$. If multiple such sequences in bin w exist, choose the one with the smallest index in the bin. If no such sequence exists, declare an encoding error.
  Figure A2 shows the construction of U^N for case 2, see the following.

Figure A2. Code-book construction for U^N in Theorem 1 for case 2.



- **(Construction of X^N)** The x^N is generated the same as that for the case 1, and it is omitted here.

Decoding:
Given a vector $y^N \in \mathcal{Y}^N$, try to find a sequence $u^N(\hat{w}, \hat{i})$ such that $(u^N(\hat{w}, \hat{i}), a^N(\hat{w}), y^N) \in T^N_{UAY}(\epsilon_3)$. If there exist sequences with the same $\hat{w}$, put out the corresponding $\hat{w}$. Otherwise, *i.e.*, if no such sequence exists or multiple sequences have different message indices, declare a decoding error.

A.2. Proof of Achievability

By using the above definitions, it is easy to verify that $\lim_{N\to\infty}\frac{\log\|\mathcal{W}\|}{N}=R$.

Then, for the two cases, note that the above encoding and decoding schemes are similar to the one used in [5]. Hence, by similar arguments as in [5], it is easy to show that $P_e\le\epsilon$ for both cases, and the proof is omitted here. It remains to show that $\lim_{N\to\infty}\Delta\ge R_e$ for the two cases, see the following.

Proof of $\lim_{N\to\infty}\Delta\ge R_e$ for case 1

$$\begin{aligned}
\lim_{N\to\infty}\Delta &= \lim_{N\to\infty}\frac{1}{N}H(W|Z^N)\\
&= \lim_{N\to\infty}\frac{1}{N}(H(W,Z^N)-H(Z^N))\\
&= \lim_{N\to\infty}\frac{1}{N}(H(W,Z^N,U^N,J)-H(J,U^N|Z^N,W)-H(Z^N))\\
&\overset{(a)}{=} \lim_{N\to\infty}\frac{1}{N}(H(Z^N|U^N)+H(U^N,J,W)-H(J,U^N|Z^N,W)-H(Z^N))\\
&\overset{(b)}{=} \lim_{N\to\infty}\frac{1}{N}(H(Z^N|U^N)+H(U^N)-H(J,U^N|Z^N,W)-H(Z^N))\\
&= \lim_{N\to\infty}\frac{1}{N}(H(U^N)-H(J,U^N|Z^N,W)-I(Z^N;U^N))\\
&= \lim_{N\to\infty}\frac{1}{N}(H(U^N)-H(J|Z^N,W)-H(U^N|Z^N,W,J)-I(Z^N;U^N))\\
&\overset{(c)}{\ge} \lim_{N\to\infty}\frac{1}{N}(H(U^N)-\log\|\mathcal{J}\|-H(U^N|Z^N,W,J)-I(Z^N;U^N))\\
&\ge \lim_{N\to\infty}\frac{1}{N}(H(U^N)-H(U^N|Y^N)-\log\|\mathcal{J}\|-H(U^N|Z^N,W,J)-I(Z^N;U^N))\\
&= \lim_{N\to\infty}\frac{1}{N}(I(Y^N;U^N)-\log\|\mathcal{J}\|-H(U^N|Z^N,W,J)-I(Z^N;U^N))\\
&\overset{(d)}{=} \lim_{N\to\infty}\frac{1}{N}(NI(Y;U)-\log\|\mathcal{J}\|-H(U^N|Z^N,W,J)-NI(Z;U))\\
&\overset{(e)}{=} \lim_{N\to\infty}\frac{1}{N}(NI(Y;U)-N\max(I(U;S|A),I(U;Z))+NI(U;Z)-N\epsilon_{3,N}-NI(Z;U))\\
&\overset{(f)}{=} I(Y;U)-\max(I(U;S|A),I(U;Z))=R_e
\end{aligned}\tag{A9}$$

where (a) is from $(W,J)\to U^N\to Z^N$, (b) is from $H(J,W|U^N)=0$, (c) is from $H(J|Z^N,W)\le H(J)\le\log\|\mathcal{J}\|$, (d) is from that S^N, U^N and X^N are i.i.d. generated random vectors, and the channels are discrete memoryless, (e) is from (A3) and (A5), and (f) is from $\epsilon_{3,N}\to 0$ as $N\to\infty$.

Thus, $\lim_{N\to\infty}\Delta\ge R_e$ for case 1 is proved.

Proof of $\lim_{N\to\infty}\Delta\ge R_e$ for case 2

$$\begin{aligned}
\lim_{N\to\infty}\Delta &= \lim_{N\to\infty}\frac{1}{N}H(W|Z^N)\\
&=^{(1)} \lim_{N\to\infty}\frac{1}{N}H(A^N|Z^N)\\
&=^{(2)} \lim_{N\to\infty}\frac{1}{N}(NH(A|Z))\\
&= H(A|Z)=R_e
\end{aligned}\tag{A10}$$

where (1) is from A^N is a function of W, and (2) is from A^N and X^N are i.i.d. generated random vectors, and the channels are discrete memoryless.

Thus, $\lim_{N\to\infty}\Delta \geq R_e$ for case 2 is proved.

The proof of Theorem 1 is completed.

B. Proof of Theorem 2

In this section, we prove Theorem 2: all the achievable (R, R_e) pairs are contained in the set $\mathcal{R}^{(no)}$. Suppose (R, R_e) is achievable, *i.e.*, for any given $\epsilon > 0$, there exists a channel encoder-decoder (N, Δ, P_e) such that

$$\lim_{N\to\infty}\frac{\log \parallel \mathcal{W} \parallel}{N} = R, \lim_{N\to\infty}\Delta \geq R_e, P_e \leq \epsilon$$

Then we will show the existence of random variables $(A, U, K, V) \to (X, S) \to Y \to Z$ such that

$$0 \leq R_e \leq R \tag{A11}$$

$$R \leq I(U;Y) - I(U;S|A) \tag{A12}$$

$$R_e \leq I(U;Y) - I(K;Z|V) \tag{A13}$$

Since W is uniformly distributed over $\mathcal{W}$, we have $H(W) = \log \parallel \mathcal{W} \parallel$. The formulas (A12) and (A13) are proved by Lemma 1, see the following.

Lemma 1 *The random vectors Y^N, Z^N and the random variables W, V, U, K, A, Y, Z of Theorem 2, satisfy:*

$$\frac{1}{N}H(W) \leq I(U;Y) - I(U;S|A) + \frac{1}{N}\delta(P_e) \tag{A14}$$

$$\frac{1}{N}H(W|Z^N) \leq I(U;Y) - I(K;Z|V) + \frac{1}{N}\delta(P_e) \tag{A15}$$

where $\delta(P_e) = h(P_e) + P_e \log(|\mathcal{W}| - 1)$. Note that $h(P_e) = -P_e \log P_e - (1 - P_e)\log(1 - P_e)$

Substituting $H(W) = \log \parallel \mathcal{W} \parallel$ and (5) into (A14) and (A15), and using the fact that $\epsilon \to 0$, the formulas (A12) and (A13) are obtained. The formula (A11) is from

$$R_e \leq \lim_{N\to\infty}\Delta = \lim_{N\to\infty}\frac{1}{N}H(W|Z^N) \leq \lim_{N\to\infty}\frac{1}{N}H(W) = R$$

It remains to prove Lemma 1, see the following.

Proof 3 (Proof of Lemma 1) *The formula (A14) follows from (A16), (A18) and (A28). The formula (A15) is from (A16), (A17), (A18), (A22), (A28) and (A29).*

<Part i> We begin with the left parts of the inequalities (A14) and (A15), see the following.

Since $W \to Y^N \to Z^N$ is a Markov chain, for the message W, we have

$$\begin{aligned}\frac{1}{N}H(W) &= \frac{1}{N}H(W|Y^N) + \frac{1}{N}I(Y^N;W)\\ &\leq^{(a)} \frac{1}{N}\delta(P_e) + \frac{1}{N}I(Y^N;W)\end{aligned} \tag{A16}$$

For the equivocation to the wiretapper, we have

$$\begin{aligned}\frac{1}{N}H(W|Z^N) &= \frac{1}{N}(H(W)-I(W;Z^N))\\ &= \frac{1}{N}(H(W)+H(W|Y^N)-H(W|Y^N)-I(W;Z^N))\\ &= \frac{1}{N}(I(W;Y^N)+H(W|Y^N)-I(W;Z^N))\\ &\leq^{(b)} \frac{1}{N}(I(W;Y^N)-I(W;Z^N)+\delta(P_e))\end{aligned} \tag{A17}$$

Note that (a) and (b) follow from Fano's inequality.

<Part ii> By using chain rule, the character $I(Y^N;W)$ in formulas (A16) and (A17) can be bounded as follows,

$$\begin{aligned}\frac{1}{N}I(Y^N;W) &= \frac{1}{N}\sum_{i=1}^{N} I(Y_i;W|Y^{i-1})\\ &=^{(1)} \frac{1}{N}\sum_{i=1}^{N}(I(Y_i;W|Y^{i-1})-I(S_i;W|S_{i+1}^N,A^N))\\ &= \frac{1}{N}\sum_{i=1}^{N}(I(Y_i;W,S_{i+1}^N,A^N|Y^{i-1})-I(Y_i;S_{i+1}^N,A^N|W,Y^{i-1})\\ &\quad -I(S_i;W,Y^{i-1}|S_{i+1}^N,A^N)+I(S_i;Y^{i-1}|W,S_{i+1}^N,A^N))\\ &=^{(2)} \frac{1}{N}\sum_{i=1}^{N}(I(Y_i;W,S_{i+1}^N,A^N|Y^{i-1})-I(S_i;W,Y^{i-1}|S_{i+1}^N,A^N))\\ &= \frac{1}{N}\sum_{i=1}^{N}(H(Y_i|Y^{i-1})-H(Y_i|Y^{i-1},W,S_{i+1}^N,A^N)-H(S_i|S_{i+1}^N,A^N)\\ &\quad +H(S_i|S_{i+1}^N,A^N,W,Y^{i-1}))\\ &\leq^{(3)} \frac{1}{N}\sum_{i=1}^{N}(H(Y_i)-H(Y_i|Y^{i-1},W,S_{i+1}^N,A^N)-H(S_i|A_i)\\ &\quad +H(S_i|S_{i+1}^N,A^N,W,Y^{i-1}))\end{aligned} \tag{A18}$$

where formula (1) follows from that $W \to A^N \to S^N$, formula (2) follows from that

$$\sum_{i=1}^{N} I(Y_i;S_{i+1}^N,A^N|W,Y^{i-1}) = \sum_{i=1}^{N} I(S_i;Y^{i-1}|W,S_{i+1}^N,A^N) \tag{A19}$$

and formula (3) follows from that $S_i \to A_i \to (S_{i+1}^N, A^{i-1}, A_{i+1}^N)$.

Proof 4 (Proof of (A19)) *The left part of (A19) can be rewritten as*

$$\begin{aligned}\sum_{i=1}^{N} I(Y_i; S_{i+1}^N, A^N | W, Y^{i-1}) &=^{(1)} \sum_{i=1}^{N} I(Y_i; S_{i+1}^N, A^N | W, Y^{i-1}, A^N) \\ &= \sum_{i=1}^{N} I(Y_i; S_{i+1}^N | W, Y^{i-1}, A^N) \\ &= \sum_{i=1}^{N}\sum_{j=i+1}^{N} I(Y_i; S_j | A^N, Y^{i-1}, W, S_{j+1}^N) \\ &= \sum_{j=1}^{N}\sum_{i=j+1}^{N} I(Y_j; S_i | A^N, Y^{j-1}, S_{i+1}^N, W) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1} I(Y_j; S_i | A^N, Y^{j-1}, S_{i+1}^N, W) \end{aligned} \tag{A20}$$

where (1) is from the fact that A^N is a deterministic function of W.

The right part of (A19) can be rewritten as

$$\sum_{i=1}^{N} I(S_i; Y^{i-1} | W, S_{i+1}^N, A^N) = \sum_{i=1}^{N}\sum_{j=1}^{i-1} I(Y_j; S_i | A^N, W, Y^{j-1}, S_{i+1}^N) \tag{A21}$$

The formula (A19) is proved by (A20) and (A21). The proof is completed.

<Part iii> Similar to (A18), the character $I(W; Z^N)$ in formula (A17) can be rewritten as follows,

$$\begin{aligned}\frac{1}{N} I(Z^N; W) &= \frac{1}{N}\sum_{i=1}^{N} I(Z_i; W | Z^{i-1}) \\ &=^{(a)} \frac{1}{N}\sum_{i=1}^{N} (I(Z_i; W | Z^{i-1}) - I(S_i; W | S_{i+1}^N, A^N)) \\ &= \frac{1}{N}\sum_{i=1}^{N} (I(Z_i; W, S_{i+1}^N, A^N | Z^{i-1}) - I(Z_i; S_{i+1}^N, A^N | W, Z^{i-1}) \\ &\quad - I(S_i; W, Z^{i-1} | S_{i+1}^N, A^N) + I(S_i; Z^{i-1} | W, S_{i+1}^N, A^N)) \\ &=^{(b)} \frac{1}{N}\sum_{i=1}^{N} (I(Z_i; W, S_{i+1}^N, A^N | Z^{i-1}) - I(S_i; W, Z^{i-1} | S_{i+1}^N, A^N)) \\ &= \frac{1}{N}\sum_{i=1}^{N} (H(Z_i | Z^{i-1}) - H(Z_i | Z^{i-1}, W, S_{i+1}^N, A^N) - H(S_i | S_{i+1}^N, A^N) \\ &\quad + H(S_i | S_{i+1}^N, A^N, W, Z^{i-1})) \\ &\geq \frac{1}{N}\sum_{i=1}^{N} (H(Z_i | Z^{i-1}) - H(Z_i | Z^{i-1}, W, S_{i+1}^N, A^N) - H(S_i | A_i) \\ &\quad + H(S_i | S_{i+1}^N, A^N, W, Z^{i-1}, Y^{i-1})) \\ &=^{(c)} \frac{1}{N}\sum_{i=1}^{N} (H(Z_i | Z^{i-1}) - H(Z_i | Z^{i-1}, W, S_{i+1}^N, A^N) - H(S_i | A_i) \\ &\quad + H(S_i | S_{i+1}^N, A^N, W, Y^{i-1})) \end{aligned} \tag{A22}$$

where formula (a) follows from that $W \to A^N \to S^N$*, formula (b) follows from that*

$$\sum_{i=1}^{N} I(Z_i; S_{i+1}^N, A^N | W, Z^{i-1}) = \sum_{i=1}^{N} I(S_i; Z^{i-1} | W, S_{i+1}^N, A^N) \tag{A23}$$

and formula (c) follows from that $Z^{i-1} \to (S_{i+1}^N, A^N, W, Y^{i-1}) \to S_i$*. Note that the proof of (A23) is similar to the proof of (A19), and therefore, it is omitted here.*

<Part iv> (single letter) To complete the proof, we introduce a random variable J*, which is independent of* W*,* A^N*,* X^N*,* S^N*,* Y^N *and* Z^N*. Furthermore,* J *is uniformly distributed over* $\{1, 2, ..., N\}$*. Define*

$$U = (W, Y^{J-1}, S_{J+1}^N, A^N, J) \tag{A24}$$

$$K = (W, Z^{J-1}, S_{J+1}^N, A^N, J) \tag{A25}$$

$$V = (Z^{J-1}, J) \tag{A26}$$

$$X = X_J, Y = Y_J, Z = Z_J, S = S_J, A = (A_J, J) \tag{A27}$$

<Part v> Then (A18) can be rewritten as

$$\begin{aligned}
\frac{1}{N} I(W; Y^N) &\leq \frac{1}{N} \sum_{i=1}^{N} (H(Y_i) - H(Y_i | Y^{i-1}, W, S_{i+1}^N, A^N) - H(S_i | A_i) + \\
&\quad H(S_i | S_{i+1}^N, A^N, W, Y^{i-1})) \\
&= \frac{1}{N} \sum_{i=1}^{N} (H(Y_i | J = i) - H(Y_i | Y^{i-1}, W, S_{i+1}^N, A^N, J = i) - H(S_i | A_i, J = i) + \\
&\quad H(S_i | S_{i+1}^N, A^N, W, Y^{i-1}, A_i, J = i)) \\
&= H(Y_J | J) - H(Y_J | Y^{J-1}, W, S_{J+1}^N, A^N, J) - H(S_J | A_J, J) + \\
&\quad H(S_J | S_{J+1}^N, A^N, W, Y^{J-1}, A_J, J) \\
&\leq H(Y_J) - H(Y_J | Y^{J-1}, W, S_{J+1}^N, A^N, J) - H(S_J | A_J, J) + \\
&\quad H(S_J | S_{J+1}^N, A^N, W, Y^{J-1}, A_J, J) \\
&= H(Y) - H(Y | U) - H(S | A) + H(S | U, A) \\
&= I(U; Y) - I(U; S | A)
\end{aligned} \tag{A28}$$

Analogously, (A22) is rewritten as follows,

$$\begin{aligned}
\frac{1}{N} I(Z^N; W) &\geq \frac{1}{N} \sum_{i=1}^{N} (H(Z_i | Z^{i-1}) - H(Z_i | Z^{i-1}, W, S_{i+1}^N, A^N) - H(S_i | A_i) + \\
&\quad H(S_i | S_{i+1}^N, A^N, W, Y^{i-1})) \\
&= \frac{1}{N} \sum_{i=1}^{N} (H(Z_i | Z^{i-1}, J = i) - H(Z_i | Z^{i-1}, W, S_{i+1}^N, A^N, J = i) - H(S_i | A_i, J = i) + \\
&\quad H(S_i | S_{i+1}^N, A^N, W, Y^{i-1}, A_i, J = i)) \\
&= H(Z_J | Z^{J-1}, J) - H(Z_J | Z^{J-1}, W, S_{J+1}^N, A^N, J) - H(S_J | A_J, J) + \\
&\quad H(S_J | S_{J+1}^N, A^N, W, Y^{J-1}, A_J, J) \\
&= H(Z | V) - H(Z | K, V) - H(S | A) + H(S | U, A) \\
&= I(Z; K | V) - I(U; S | A)
\end{aligned} \tag{A29}$$

Substituting (A28), (A29) into (A16) and (A17), Lemma 1 is proved.

In addition, by using the definitions of U, K, V, Y and Z (see (A24), (A25), (A26) and (A27), note that V is a part of K), and observing that $Z^{J-1} \to (Y^{J-1}, W, S_{J+1}^N, A^N, J) \to Y_J \to Z_J$ is a Markov chain, it is easy to check that the Markov chain $V \to K \to U \to Y \to Z$ holds.

The proof of Theorem 2 is completed.

C. Size Constraint of The Random Variables in Theorem 1

By using the support lemma (see [13], p.310), it suffices to show that the random variable U can be replaced by new one, preserving the Markovity $(U, A) \to (X, S) \to Y \to Z$ and the mutual information $I(U;Z)$, $I(U;Y)$, $I(U;S|A)$, and furthermore, the range of the new U satisfies: $\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|+2$. The proof is in the reminder of this section.

Let

$$\bar{p} = p_{XSA}(x, s, a) \tag{A30}$$

Define the following continuous scalar functions of $\bar{p}$:

$$f_{XSA}(\bar{p}) = p_{XSA}(x, s, a), f_Y(\bar{p}) = H(Y), f_Z(\bar{p}) = H(Z), f_{S|A}(\bar{p}) = H(S|A)$$

Since there are $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|-1$ functions of $f_{XSA}(\bar{p})$, the total number of the continuous scalar functions of $\bar{p}$ is $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$+2.

Let $\bar{p}_{XSA|U} = Pr\{X = x, S = s, A = a|U = u\}$. With these distributions $\bar{p}_{XSA|U} = Pr\{X = x, S = s, A = a|U = u\}$, we have

$$p_{XSA}(x, s, a) = \sum_{u\in\mathcal{U}} p(U = u) f_{XSA}(\bar{p}_{XSA|U}) \tag{A31}$$

$$I(U;Z) = f_Z(\bar{p}) - \sum_{u\in\mathcal{U}} p(U = u) f_Z(\bar{p}_{XSA|U}) \tag{A32}$$

$$I(U;S|A) = f_{S|A}(\bar{p}) - \sum_{u\in\mathcal{U}} p(U = u) f_{S|A}(\bar{p}_{XSA|U}) \tag{A33}$$

$$H(Y|U) = \sum_{u\in\mathcal{U}} p(U = u) f_Y(\bar{p}_{XSA|U}) \tag{A34}$$

According to the support lemma ([13], p.310), the random variable U can be replaced by new ones such that the new U takes at most $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| + 2$ different values and the expressions (A31), (A32), (A33) and (A34) are preserved.

D. Size Constraint of The Random Variables in Theorem 2

By using the support lemma (see [13], p.310), it suffices to show that the random variables U, V and K can be replaced by new ones, preserving the Markovities $(U, K, A, V) \to (X, S) \to Y \to Z$, $V \to K \to Y \to Z$ and the mutual information $I(U;Y)$, $I(K;Z|V)$, $I(U;S|A)$, and furthermore, the ranges of the new U, V and K satisfy: $\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| + 1$, $\|\mathcal{V}\| \leq \|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$, $\|\mathcal{K}\| \leq \|\mathcal{X}\|^2\|\mathcal{S}\|^2\|\mathcal{A}\|^2$. The proof is in the reminder of this section.

- **(Proof of $\|\mathcal{U}\| \leq \|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| + 1$)**
  Let
  $$\bar{p} = p_{XSA}(x, s, a) \tag{A35}$$
  Define the following continuous scalar functions of $\bar{p}$:
  $$f_{XSA}(\bar{p}) = p_{XSA}(x, s, a), f_Y(\bar{p}) = H(Y), f_{S|A}(\bar{p}) = H(S|A)$$
  Since there are $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| - 1$ functions of $f_{XSA}(\bar{p})$, the total number of the continuous scalar functions of $\bar{p}$ is $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$+1.
  Let $\bar{p}_{XSA|U} = Pr\{X = x, S = s, A = a|U = u\}$. With these distributions $\bar{p}_{XSA|U} = Pr\{X = x, S = s, A = a|U = u\}$, we have
  $$p_{XSA}(x, s, a) = \sum_{u \in \mathcal{U}} p(U = u) f_{XSA}(\bar{p}_{XSA|U}) \tag{A36}$$
  $$I(U; S|A) = f_{S|A}(\bar{p}) - \sum_{u \in \mathcal{U}} p(U = u) f_{S|A}(\bar{p}_{XSA|U}) \tag{A37}$$
  $$H(Y|U) = \sum_{u \in \mathcal{U}} p(U = u) f_Y(\bar{p}_{XSA|U}) \tag{A38}$$
  According to the support lemma ([13], p.310), the random variable U can be replaced by new ones such that the new U takes at most $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| + 1$ different values and the expressions (A36), (A37) and (A38) are preserved.
- **(Proof of $\|\mathcal{V}\| \leq \|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$)**
  Let
  $$\bar{p} = p_{XSA}(x, s, a) \tag{A39}$$
  Define the following continuous scalar functions of $\bar{p}$:
  $$f_{XSA}(\bar{p}) = p_{XSA}(x, s, a), f_Z(\bar{p}) = H(Z)$$
  Since there are $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| - 1$ functions of $f_{XSA}(\bar{p})$, the total number of the continuous scalar functions of $\bar{p}$ is $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$.
  Let $\bar{p}_{XSA|V} = Pr\{X = x, S = s, A = a|V = v\}$. With these distributions $\bar{p}_{XSA|V} = Pr\{X = x, S = s, A = a|V = v\}$, we have
  $$p_{XSA}(x, s, a) = \sum_{v \in \mathcal{V}} p(V = v) f_{XSA}(\bar{p}_{XSA|V}) \tag{A40}$$
  $$H(Z|V) = \sum_{v \in \mathcal{V}} p(V = v) f_Z(\bar{p}_{XSA|V}) \tag{A41}$$
  According to the support lemma ([13], p.310), the random variable V can be replaced by new ones such that the new V takes at most $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$ different values and the expressions (A40) and (A41) are preserved.
- **(Proof of $\|\mathcal{K}\| \leq \|\mathcal{X}\|^2\|\mathcal{S}\|^2\|\mathcal{A}\|^2$)**
  Once the alphabet of V is fixed, we apply similar arguments to bound the alphabet of K, see the following. Define $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$ continuous scalar functions of $\bar{p}_{XSA}$:
  $$f_{XSA}(\bar{p}_{XSA}) = p_{XSA}(x, s, a), f_Z(\bar{p}_{XSA}) = H(Z)$$

where of the functions $f_{XSA}(\bar{p}_{XSA})$, only $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\| - 1$ are to be considered.
For every fixed v, let $\bar{p}_{XSA|K,V} = Pr\{X = x, S = s, A = a|K = k, V = v\}$. With these distributions $\bar{p}_{XSA|K,V}$, we have

$$Pr\{X = x, S = s, A = a|V = v\} = \sum_{k\in\mathcal{K}} Pr\{K = k|V = v\} f_{XSA}(\bar{p}_{XSA|K,V}) \tag{A42}$$

$$I(K; Z|V = v) = H(Z|V = v) - \sum_{k\in\mathcal{K}} f_Z(\bar{p}_{XSA|K,V}) Pr\{K = k|V = v\} \tag{A43}$$

By the support lemma ([13], p.310), for every fixed v, the size of the alphabet of the random variable K can not be larger than $\|\mathcal{X}\|\|\mathcal{S}\|\|\mathcal{A}\|$, and therefore, $\|\mathcal{K}\| \leq \|\mathcal{X}\|^2\|\mathcal{S}\|^2\|\mathcal{A}\|^2$ is proved.

E. Proof of Theorem 3

In this section, we will show that any pair $(R, R_e) \in \mathcal{R}^{ci}$ is achievable. Wyner's random binning technique is used in the construction of the code-book.

Now the remainder of this section is organized as follows. The code construction is in Subsection E.1. The proof of achievability is given in Subsection E.2.

E.1. Code Construction

Since $R_e \leq I(U; Y) - I(U; Z)$, $R_e \leq H(A|Z)$ and $R_e \leq R \leq I(U; Y)$, it is sufficient to show that the pair $(R, R_e = \min\{I(U; Y) - I(U; Z), H(A|Z)\}$ is achievable, and note that this implies that $R \geq R_e = \min\{I(U; Y) - I(U; Z), H(A|Z)\}$.

The construction of the code and the proof of achievability are considered into two cases:

- **(Case 1)** If $H(A|Z) \geq I(U; Y) - I(U; Z)$, Wyner's random binning technique [6] is used in the construction of the code-book.
- **(Case 2)** If $H(A|Z) \leq I(U; Y) - I(U; Z)$, Shannon's strategy [1] is used in the construction of the code-book.

- **(Code construction for case 1)**
  Given a pair (R, R_e), choose a joint probability mass function $p_{U,A,S,X,Y,Z}(u, a, s, x, y, z)$ such that

$$0 \leq R_e \leq R$$

$$R \leq I(U; Y)$$

$$R_e = I(U; Y) - I(U; Z)$$

  The message set $\mathcal{W}$ satisfies the following condition:

$$\lim_{N\to\infty} \frac{1}{N} \log \| \mathcal{W} \| = R = I(U; Y) - \gamma \tag{A44}$$

  where γ is a fixed positive real numbers and

$$0 \leq \gamma \leq^{(a)} I(U; Z) \tag{A45}$$

  Note that (a) is from $R \geq R_e = I(U; Y) - I(U; Z)$ and (A44). Let $\mathcal{W} = \{1, 2, ..., 2^{NR}\}$.
  Code-book generation:

– **(Construction of A^N)**
Generate 2^{NR} i.i.d. sequences a^N, according to the probability mass function $p_A(a)$. Index each sequence by $i \in \{1, 2, ..., 2^{NR}\}$. For a given message w ($w \in \mathcal{W}$), choose a corresponding $a^N(w)$ as the output of the action encoder.

– **(Construction of U^N)**
For the transmitted action sequence $a^N(w)$, generate $2^{N(I(U;Y)-\epsilon_{2,N})}$ ($\epsilon_{2,N} \to 0$ as $N \to \infty$) i.i.d. sequences u^N, according to the probability mass function $p_{U|A}(u_i|a_i(w))$. Distribute these sequences at random into $2^{NR} = 2^{N(I(U;Y)-\gamma)}$ bins such that each bin contains $2^{N(\gamma-\epsilon_{2,N})}$ sequences. Index each bin by $i \in \{1, 2, ..., 2^{NR}\}$.
Here note that the number of the sequences in every bin is upper bounded as follows.

$$\gamma - \epsilon_{2,N} \leq^{(a)} I(U;Z) - \epsilon_{2,N} \tag{A46}$$

where (a) is from (A45). This implies that

$$\lim_{N\to\infty} H(U^N|W, Z^N) = 0 \tag{A47}$$

Note that (A47) can be proved by using Fano's inequality and (A46).
For a given message w ($w \in \mathcal{W}$), randomly choose a sequence $u^N(w, i^*)$ in bin w as the realization of U^N.
Let s^N be the state sequence generated in response to the action sequence $a^N(w)$.

– **(Construction of X^N)** The x^N is generated according to a new discrete memoryless channel (DMC) with inputs u^N, s^N, and output x^N. The transition probability of this new DMC is $p_{X|U,S}(x|u, s)$, which is obtained from the joint probability mass function $p_{U,A,S,X,Y,Z}(u, a, s, x, y, z)$. The probability
$p_{X^N|U^N,S^N}(x^N|u^N, s^N)$ is calculated as follows.

$$p_{X^N|U^N,S^N}(x^N|u^N, s^N) = \prod_{i=1}^{N} p_{X|U,S}(x_i|u_i, s_i) \tag{A48}$$

Decoding:
Given a vector $y^N \in \mathcal{Y}^N$, try to find a sequence $u^N(\hat{w}, \hat{i})$ such that $(u^N(\hat{w}, \hat{i}), a^N(\hat{w}), y^N) \in T^N_{UAY}(\epsilon_3)$. If there exist sequences with the same $\hat{w}$, put out the corresponding $\hat{w}$. Otherwise, *i.e.*, if no such sequence exists or multiple sequences have different message indices, declare a decoding error.

- **(Code construction for case 2)**
Given a pair (R, R_e), choose a joint probability mass function $p_{U,A,S,X,Y,Z}(u, a, s, x, y, z)$ such that

$$0 \leq R_e \leq R$$

$$R \leq I(U;Y)$$

$$R_e = H(A|Z)$$

The message set $\mathcal{W}$ satisfies the following condition:

$$\lim_{N\to\infty} \frac{1}{N} \log \parallel \mathcal{W} \parallel = R = I(U;Y) - \gamma_1 \tag{A49}$$

where γ_1 is a fixed positive real numbers and

$$0 \leq \gamma_1 \leq^{(b)} I(U;Y) - H(A|Z) \tag{A50}$$

Note that (b) is from $R \geq R_e = H(A|Z)$ and (A49). Let $\mathcal{W} = \{1, 2, ..., 2^{NR}\}$.

Code-book generation:

- **(Construction of A^N)**
  Generate 2^{NR} i.i.d. sequences a^N, according to the probability mass function $p_A(a)$. Index each sequence by $i \in \{1, 2, ..., 2^{NR}\}$. For a given message w ($w \in \mathcal{W}$), choose a corresponding $a^N(w)$ as the output of the action encoder.
- **(Construction of U^N)**
  For the transmitted action sequence $a^N(w)$, generate 2^{NR} i.i.d. sequences u^N, according to the probability mass function $p_{U|A}(u_i|a_i(w))$. Index each u^N by $i \in \{1, 2, ..., 2^{NR}\}$.
  For a given message w ($w \in \mathcal{W}$), choose a sequence $u^N(w)$ as the realization of U^N.
  Let s^N be the state sequence generated in response to the action sequence $a^N(w)$.
- **(Construction of X^N)** The x^N is generated the same as that for the case 1, and it is omitted here.

Decoding:

Given a vector $y^N \in \mathcal{Y}^N$, try to find a sequence $u^N(\hat{w})$ such that $(u^N(\hat{w}), a^N(\hat{w}), y^N) \in T^N_{UAY}(\epsilon_3)$. If there exist sequences with the same $\hat{w}$, put out the corresponding $\hat{w}$. Otherwise, *i.e.*, if no such sequence exists or multiple sequences have different message indices, declare a decoding error.

E.2. Proof of Achievability

By using the above definitions, it is easy to verify that $\lim_{N\to\infty} \frac{\log\|\mathcal{W}\|}{N} = R$.

Then, for the two cases, note that the above encoding and decoding schemes are similar to the one used in [5]. Hence, by similar arguments as in [5], it is easy to show that $P_e \leq \epsilon$ for both cases, and the proof is omitted here. It remains to show that $\lim_{N\to\infty} \Delta \geq R_e$ for the two cases, see the following.

Proof of $\lim_{N\to\infty} \Delta \geq R_e$ for case 1

$$\begin{aligned}
\lim_{N\to\infty} \Delta &= \lim_{N\to\infty} \frac{1}{N} H(W|Z^N) = \lim_{N\to\infty} \frac{1}{N}(H(W, Z^N) - H(Z^N)) \\
&= \lim_{N\to\infty} \frac{1}{N}(H(W, Z^N, U^N) - H(U^N|Z^N, W) - H(Z^N)) \\
&\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N}(H(Z^N|U^N) + H(U^N, W) - H(U^N|Z^N, W) - H(Z^N)) \\
&\overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N}(H(Z^N|U^N) + H(U^N) - H(U^N|Z^N, W) - H(Z^N)) \\
&= \lim_{N\to\infty} \frac{1}{N}(H(U^N) - H(U^N|Z^N, W) - I(Z^N; U^N)) \\
&\geq \lim_{N\to\infty} \frac{1}{N}(H(U^N) - H(U^N|Y^N) - H(U^N|Z^N, W) - I(Z^N; U^N)) \\
&= \lim_{N\to\infty} \frac{1}{N}(I(U^N; Y^N) - H(U^N|Z^N, W) - I(Z^N; U^N)) \\
&=^{(c)} \lim_{N\to\infty} \frac{1}{N}(NI(U;Y) - NI(U;Z)) \\
&= I(U;Y) - I(U;Z) = R_e,
\end{aligned} \tag{A51}$$

where (a) is from $W \to U^N \to Z^N$, (b) is from $H(W|U^N) = 0$, (c) is from that S^N, U^N and X^N are i.i.d. generated random vectors, the channels are discrete memoryless and (A47).

Thus, $\lim_{N\to\infty} \Delta \geq R_e$ for case 1 is proved.

Proof of $\lim_{N\to\infty} \Delta \geq R_e$ for case 2

$$\begin{aligned} \lim_{N\to\infty} \Delta &= \lim_{N\to\infty} \frac{1}{N} H(W|Z^N) \\ &=^{(1)} \lim_{N\to\infty} \frac{1}{N} H(A^N|Z^N) \\ &=^{(2)} \lim_{N\to\infty} \frac{1}{N} (NH(A|Z)) \\ &= H(A|Z) = R_e, \end{aligned} \tag{A52}$$

where (1) is from A^N is a function of W, and (2) is from A^N and X^N are i.i.d. generated random vectors, and the channels are discrete memoryless.

Thus, $\lim_{N\to\infty} \Delta \geq R_e$ for case 2 is proved.

The proof of Theorem 3 is completed.

F. Proof of Theorem 4

In this section, we prove Theorem 4: all the achievable (R, R_e) pairs are contained in the set $\mathcal{R}^{(co)}$. Suppose (R, R_e) is achievable, *i.e.*, for any given $\epsilon > 0$, there exists a channel encoder-decoder (N, Δ, P_e) such that

$$\lim_{N\to\infty} \frac{\log \parallel \mathcal{W} \parallel}{N} = R, \lim_{N\to\infty} \Delta \geq R_e, P_e \leq \epsilon.$$

Then we will show the existence of random variables $(A, U, K, V) \to (X, S) \to Y \to Z$ such that

$$0 \leq R_e \leq R \tag{A53}$$

$$R \leq I(U;Y) \tag{A54}$$

$$R_e \leq I(U;Y) - I(K;Z|V) \tag{A55}$$

Since W is uniformly distributed over $\mathcal{W}$, we have $H(W) = \log \parallel \mathcal{W} \parallel$. The formulas (A54) and (A55) are proved by Lemma 2, see the following.

Lemma 2 *The random vectors Y^N, Z^N and the random variables W, V, U, K, A, Y, Z of Theorem 4, satisfy:*

$$\frac{1}{N} H(W) \leq I(U;Y) + \frac{1}{N} \delta(P_e), \tag{A56}$$

$$\frac{1}{N} H(W|Z^N) \leq I(U;Y) - I(K;Z|V) + \frac{1}{N} \delta(P_e), \tag{A57}$$

where $\delta(P_e) = h(P_e) + P_e \log(|\mathcal{W}| - 1)$. Note that $h(P_e) = -P_e \log P_e - (1 - P_e) \log(1 - P_e)$.

Substituting $H(W) = \log \parallel \mathcal{W} \parallel$ and (5) into (A56) and (A57), and using the fact that $\epsilon \to 0$, the formulas (A54) and (A55) are obtained. The formula (A53) is from

$$R_e \leq \lim_{N\to\infty} \Delta = \lim_{N\to\infty} \frac{1}{N} H(W|Z^N) \leq \lim_{N\to\infty} \frac{1}{N} H(W) = R.$$

It remains to prove Lemma 2, see the following.

Proof 5 (Proof of Lemma 2) *The formula (A56) follows from (A58), (A60) and (A66). The formula (A57) is from (A58), (A59), (A60), (A61), (A66) and (A67).*

<Part i> We begin with the left parts of the inequalities (A56) and (A57), see the following.

Since $W \to Y^N \to Z^N$ is a Markov chain, for the message W, we have

$$\begin{aligned} \frac{1}{N}H(W) &= \frac{1}{N}H(W|Y^N) + \frac{1}{N}I(Y^N;W) \\ &\leq^{(a)} \frac{1}{N}\delta(P_e) + \frac{1}{N}I(Y^N;W). \end{aligned} \tag{A58}$$

For the equivocation to the wiretapper, we have

$$\begin{aligned} \frac{1}{N}H(W|Z^N) &= \frac{1}{N}(H(W) - I(W;Z^N)) \\ &= \frac{1}{N}(H(W) + H(W|Y^N) - H(W|Y^N) - I(W;Z^N)) \\ &= \frac{1}{N}(I(W;Y^N) + H(W|Y^N) - I(W;Z^N)) \\ &\leq^{(b)} \frac{1}{N}(I(W;Y^N) - I(W;Z^N) + \delta(P_e)). \end{aligned} \tag{A59}$$

Note that (a) and (b) follow from Fano's inequality.

<Part ii> By using chain rule, the character $I(Y^N;W)$ in formulas (A58) and (A59) can be bounded as follows,

$$\begin{aligned} \frac{1}{N}I(Y^N;W) &= \frac{1}{N}\sum_{i=1}^{N} I(Y_i;W|Y^{i-1}) \\ &\leq \frac{1}{N}\sum_{i=1}^{N} I(Y_i;W,Y^{i-1}) \\ &\leq \frac{1}{N}\sum_{i=1}^{N} I(Y_i;W,Y^{i-1},S^{i-1}). \end{aligned} \tag{A60}$$

<Part iii> Similar to (A60), the character $I(W;Z^N)$ in formula (A59) can be rewritten as follows,

$$\begin{aligned} \frac{1}{N}I(Z^N;W) &= \frac{1}{N}\sum_{i=1}^{N} I(Z_i;W|Z^{i-1}) \\ &= \frac{1}{N}\sum_{i=1}^{N}(H(Z_i|Z^{i-1}) - H(Z_i|Z^{i-1},W)). \end{aligned} \tag{A61}$$

<Part iv> (single letter) To complete the proof, we introduce a random variable J, which is independent of W, A^N, X^N, S^N, Y^N and Z^N. Furthermore, J is uniformly distributed over $\{1,2,...,N\}$. Define

$$U = (W, Y^{J-1}, S^{J-1}, J) \tag{A62}$$

$$K = (W, Z^{J-1}, J) \tag{A63}$$

$$V = (Z^{J-1}, J) \tag{A64}$$

$$X = X_J, Y = Y_J, Z = Z_J, S = S_J, A = A_J \tag{A65}$$

<Part v> Then (A60) can be rewritten as

$$\begin{aligned}\frac{1}{N}I(W;Y^N) &\leq \frac{1}{N}\sum_{i=1}^{N} I(Y_i;W,Y^{i-1},S^{i-1}) \\ &= \frac{1}{N}\sum_{i=1}^{N} I(Y_i;W,Y^{i-1},S^{i-1}|J=i) \\ &= I(Y_J;W,Y^{J-1},S^{J-1}|J) \\ &\leq I(Y_J;W,Y^{J-1},S^{J-1},J) \\ &= I(U;Y) \end{aligned} \tag{A66}$$

Analogously, (A61) is rewritten as follows,

$$\begin{aligned}\frac{1}{N}I(Z^N;W) &= \frac{1}{N}\sum_{i=1}^{N}(H(Z_i|Z^{i-1}) - H(Z_i|Z^{i-1},W)) \\ &= \frac{1}{N}\sum_{i=1}^{N}(H(Z_i|Z^{i-1},J=i) - H(Z_i|Z^{i-1},W,J=i)) \\ &= H(Z_J|Z^{J-1},J) - H(Z_J|Z^{J-1},W,J) \\ &= H(Z|V) - H(Z|K,V) \\ &= I(Z;K|V) \end{aligned} \tag{A67}$$

Substituting (A66), (A67) into (A58) and (A59), Lemma 2 is proved.

In addition, by using the definitions of U, K, V, Y and Z (see (A62), (A63), (A64) and (A65), note that V is a part of K), and observing that $Z^{J-1} \to (Y^{J-1},W,S^{J-1},J) \to Y_J \to Z_J$ and $(W,Y^{J-1},S^{J-1},J) \to A_J \to S_J$ are two Markov chains, it is easy to check that the Markov chains $V \to K \to U \to Y \to Z$ and $U \to A \to S$ hold.

The proof of Theorem 4 is completed.