Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 131063, 19 pages
http://dx.doi.org/10.1155/2013/131063

*Research Article*

# Wiretap Channel with Action-Dependent States and Rate-Limited Feedback

**Xinxing Yin, Xiao Chen, and Zhi Xue**

*Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

Correspondence should be addressed to Xinxing Yin; yinxinxing@sjtu.edu.cn

Received 5 September 2013; Revised 1 November 2013; Accepted 2 November 2013

Academic Editor: Hamid Reza Karimi



We introduce the wiretap channel with action-dependent states and rate-limited feedback. In the new model, the state sequence is dependent on the action sequence which is selected according to the message, and a secure rate-limited feedback link is shared between the transmitter and the receiver. We obtain the capacity-equivocation region and secrecy capacity of such a channel both for the case where the channel inputs depend noncausally on the state sequence and the case where they are restricted to causal dependence. We construct the capacity-achieving coding schemes utilizing Wyner's random binning, Gel'fand and Pinsker's coding technique, and rate splitting. Furthermore, we compare our results with the existing approaches without feedback, with noiseless feedback, and without action-dependent states. The simulation results show that the secrecy capacity of our model is bigger than that of the first two existed approaches. Besides, it is also shown that, by taking actions to affect the channel states, we guarantee the data integrity of the message transmitted in the two-stage communication systems although the tolerable overhead of transmission time is brought.

## 1. Introduction

The framework of channels with action-dependent states was first introduced by Weissman [1] to model scenarios in which transmission took place in two successive phases. In the first phase, a message-dependent action sequence was selected by the encoder. The action sequence affected the formation of the channel states. In the second phase, the transmitter began to send the message to the receiver in the presence of the action-dependent states. In [1], the capacity of such a channel both for the case where the channel inputs were allowed to depend noncausally on the channel states and the case where they were restricted to causal dependence was obtained. After the publication of Weissman's work, a number of extensions of the results in [1] have been reported; see [2–6].

However, the above scenarios considered no security constraints which are essential in many communication systems. For instance, the broadcast nature of wireless networks gives rise to the hidden danger of information leakage to the malicious receiver when broadcasting sensitive data and acquiring channel state information. The secure communication for the wiretap channel was first studied by Wyner [7]. In his model, the transmitter aimed to send a confidential message to the receiver through a noisy discrete memoryless channel (DMC) and keep the wiretapper as ignorant of the message as possible. The wiretapper observed a degraded version of the legitimate receiver's observation through another DMC. Equivocation was introduced in [7] to measure wiretapper's uncertainty of the confidential message. Wyner characterized the secrecy capacity, that is, the best transmission rate under perfect secrecy, as the difference between the capacity of the main channel and the wiretap channel. This model was further explored by many researchers; see [8–14]. Among them, Dai et al. studied the wiretap channel with action-dependent states [13] and gave the lower and upper bounds on the *capacity-equivocation region*. The *capacity-equivocation region* is the set of all the achievable rate pairs $(R, R_e)$, where $R$ and $R_e$ are the rates of the confidential message and wiretapper's equivocation about the message.

Note that the action-dependent channel models [1–6] as well as the wiretap channel models [7–14] only dealt with one-way communication. However, many systems involve two-way communication. For example, feedback links are usually seen in satellite communication, telephone connections, and

wireless sensor networks. To investigate the effects of the feedback on secrecy capacity, Ahlswede and Cai first studied the wiretap channel with noiseless causal feedback [15], where the channel output symbol is fed back to the transmitter and used as a secret key. Then, Ardestanizadeh et al. studied the problem of secure communication over a wiretap channel with a rate-limited feedback link [16]. The main contribution of [16] was that the upper and lower bounds on the secrecy capacity were obtained as a function of the secure feedback rate $R_f$. Moreover, the recursive argument was introduced to find the single-letter characterization of the upper bound and claimed to be a powerful tool for similar problems [16]. Besides, Yin et al. [17] studied the discrete memoryless broadcast channels with noiseless feedback based on [15, 16]. The channel model in [17] did not consider channel state information. Recently, Dai et al. studied the wiretap channel with action-dependent states and noiseless feedback [18]. In Dai's model, similar to [15], the output symbol received at the receiver was fed back securely to the transmitter and used as a shared key between the transmitter and the receiver. However, in [15, 18], only part of the feedback contributed to the secure communication and the wiretapper was provided a chance to get the secret message by guessing the legitimate receiver's channel output with its own channel observation. Then, it is natural to ask whether the feedback can be used more efficiently and securely.

Inspired by [16, 18], this paper studies a new model, that is, wiretap channel with action-dependent states and rate-limited feedback; see Figure 1. In the model, the feedback is independent of the channel output and sent to the transmitter through a secure feedback link of rate $R_f$. The formation of the channel states is affected by the message-dependent action sequence. This model can provide insights into the value of two-way two-phase interactions in communication systems. Note that since the feedback symbol is independent of the channel output, the wiretapper attempt to get the secret key by guessing the legitimate receiver's channel output is in vain. The contributions of this work are summarized as follows.

(i) The capacity-equivocation region of the channel model in Figure 1 is obtained both for the case where the inputs of the main channel depend noncausally on the channel states and the case where they depend causally on the channel states. The capacity-equivocation region is presented in Section 2. Besides, the secrecy capacity of the channel model in Figure 1 is also got. We calculate the secrecy capacity of a binary example in Section 3.

(ii) To achieve the secrecy capacity, we construct several coding schemes and evaluate their reliability and security using the techniques of Wyner's random binning, Gel'fand and Pinsker's coding, and rate splitting. The coding schemes are presented in the Appendices.

(iii) The rate-limited feedback is added in our model. The secrecy capacity of our approach is bigger than that of the model without feedback. This indicates that feedback is useful for increasing the secrecy capacity. The corresponding simulation is presented in Section 3.

(iv) The capacity-equivocation region of the wiretap channel with action-dependent states and noiseless feedback [18] is included in our results by setting the feedback rate to be a specific value. We find that the secrecy capacity of our model is bigger than that of [18]. Specifically, when the wiretap channel is in the "best" condition, it is unable to transmit message securely using the approach in [18], while our approach can still work. Section 3 discusses the result in detail.

(v) The (degraded) wiretap channel with secure rate-limited feedback [16] is a special case of our model. The secrecy capacity of [16] can be obtained from our results without considering the action-dependent states. Our approach can guarantee data integrity in the two-stage systems. However, data integrity cannot be guaranteed by using the approach in [16] where no actions were taken. This result is illustrated in Section 3.

The remainder of the paper is organized as follows. Section 2 presents our new model and two theorems. Section 3 gives the secrecy capacity of our model and compares it with the existing channel models through simulation. We conclude in Section 4 with a summary of the whole work and some future directions.

## 2. Channel Models and Main Results

In this section, the notations of characters and variables are given in Section 2.1. The channel model is described in Section 2.2. The main results, that is, the capacity-equivocation region of the model in Figure 1 with causal and noncausal states, are presented in Section 2.3.

*2.1. Notations.* Throughout this paper, we use calligraphic letters, for example, $\mathcal{X}$, $\mathcal{Y}$, to denote the finite sets and $\|\mathcal{X}\|$ to denote the cardinality of the set $\mathcal{X}$. Uppercase letters, for example, $X$, $Y$, are used to denote random variables taking values from finite sets, for example, $\mathcal{X}$, $\mathcal{Y}$. The value of a random variable $X$ is denoted by the lowercase letter $x$. We use $Z_i^j$ to denote the $(j-i+1)$-vectors $(Z_i, Z_{i+1}, \ldots, Z_j)$ of random variables for $1 \le i \le j$, and we will always drop the subscript when $i = 1$. Moreover, we use $X \sim p(x)$ to denote the probability mass function of the random variable $X$. For $X \sim p(x)$ and $0 \le \epsilon \le 1$, the set of the typical $N$-sequences $x^N$ is defined as $\mathcal{T}_X^N(\epsilon) = \{x^N : |\pi(x \mid x^N) - p(x)| \le \epsilon p(x)$ for all $x \in \mathcal{X}\}$, where $\pi(x \mid x^N)$ denotes the frequency of occurrences of letter $x$ in the sequence $x^N$. (For more details about typical sequences, please refer to [19, Chap. 2].) The set of the conditional typical sequences, for example, $\mathcal{T}_{Y|X}^N(\epsilon)$, follows similarly. In this paper, it is assumed that the base of the log function is 2.

*2.2. Wiretap Channel with Action-Dependent States and Rate-Limited Feedback.* We consider the wiretap channel with

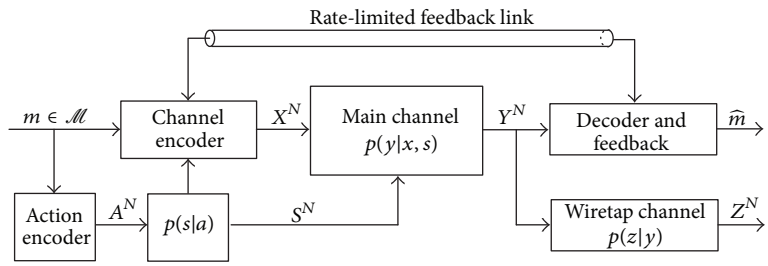


Figure 1: Wiretap channel with action-dependent states and rate-limited feedback.

action-dependent states and rate-limited feedback, as shown in Figure 1. The transmitter aims to convey a confidential message $M$ which is uniformly distributed over $\mathscr{M}$ to the legitimate receiver whose observation is $Y^N \in \mathscr{Y}^N$. The confidential message should be kept secret from the wiretapper as much as possible. We use equivocation at the wiretapper to characterize the secrecy of the confidential message. Given the message $M$, an action sequence $A^N(M) \in \mathscr{A}^N$ is selected. The channel state sequence $S^N \in \mathscr{S}^N$ is generated in response to the action sequence and accessible to the channel encoder. To enhance the secrecy of the communication, the receiver feeds back symbols $K \in \mathscr{K}$ to the encoder over a feedback link of rate $R_f$. The feedback symbol is assumed to be independent of the channel output $Y^N$ and kept secret from the wiretapper. Then, the input sequence of the main channel $X^N \in \mathscr{X}^N$ is generated based on the message $M$, the channel state sequence $S^N$, and the feedback symbol $K$. The output sequence of the main channel $Y^N$ is distributed as $p(y^N \mid s^N, x^N) = \prod_{i=1}^N p(y_i \mid s_i, x_i)$. With the received $Y^N$, the decoder outputs the decoded message $\widehat{M} \in \mathscr{M}$. The wiretap channel is also a DMC with transition probability $p(z^N \mid y^N) = \prod_{i=1}^N p(z_i \mid y_i)$, where $Z^N \in \mathscr{Z}^N$ is the observation of the wiretapper. Note that the wiretap channel is assumed to be degraded from the main channel; that is, $X \to Y \to Z$ form a Markov chain. More precisely, we define the $(2^{NR}, 2^{NR_f}, N)$ code in Definition 1.

*Definition 1.* The $(2^{NR}, 2^{NR_f}, N)$ code for the wiretap channel with action-dependent states and rate-limited feedback is defined as follows.

(i) The feedback alphabet $\mathscr{K}$ satisfies $\lim_{N\to\infty} (\log \|\mathscr{K}\|/N) \le R_f$. The feedback symbol is generated independently of the channel output symbols and uniformly distributed over $\mathscr{K}$.

(ii) The message $M$ is uniformly distributed over $\mathscr{M}$. The action sequence $A^N(M) \in \mathscr{A}^N$ is selected from a deterministic mapping $g : \mathscr{M} \to \mathscr{A}^N$. The state sequence is generated as the output of a memoryless channel $p(s^N \mid a^N) = \prod_{i=1}^N p(s_i \mid a_i)$ whose input is $A^N$.

(iii) The stochastic channel encoder $\varphi$ is specified by a matrix of conditional probability distributions $\varphi(x^N \mid m, s^N, k^N)$, where $m \in \mathscr{M}$, $s^N \in \mathscr{S}^N$, $k^N \in \mathscr{K}^N$, and $\sum_{x^N} \varphi(x^N \mid m, s^N, k^N) = 1$. Note that $\varphi(x^N \mid m, s^N, k^N)$ is the probability that the message $m$, the state sequence $s^N$, and the feedback $k^N$ are encoded as the channel input $x^N$. When the state sequence $s^N$ is known causally to the channel encoder, the channel encoder at time $i$ is $\varphi_i(x_i \mid m, s^i, k_i)$, where $x_i$ is the output of the channel encoder at time $i$, $k_i$ is the feedback symbol at time $i$, and $s^i = (s_1, s_2, \ldots, s_i)$ is the channel states before time $i$. When the channel encoder knows the state sequence $s^N$ in a noncausal manner, the channel encoder at time $i$ is $\varphi_i(x_i \mid m, s^N, k_i)$. For the causal manner, Shannon strategy will be used in the encoding scheme (see Appendix A). For the noncausal manner, Gel'fand and Pinsker's coding technique [20] will be used (see Appendix C).

(iv) The decoder is a mapping $\psi : \mathscr{Y}^N \to \mathscr{M}$. The input of the decoder is $Y^N$ and the output is $\widehat{M}$. The decoding error probability is defined as $P_e = \Pr\{\psi(Y^N) \neq M\}$.

(v) The equivocation of the message at the wiretapper is defined as

$$\Delta = \frac{1}{N} H\left(M \mid Z^N\right). \tag{1}$$

*Definition 2.* A rate pair $(R, R_e)$ is said to be achievable for the model in Figure 1 if there exists a $(2^{NR}, 2^{NR_f}, N)$ code defined in Definition 1, such that

$$\begin{gathered} \lim_{N\to\infty} \frac{\log \|\mathscr{M}\|}{N} = R; \\ \lim_{N\to\infty} \frac{\log \|\mathscr{K}\|}{N} \le R_f; \\ \lim_{N\to\infty} \Delta \ge R_e; \\ P_e \le \epsilon, \end{gathered} \tag{2}$$

where $\epsilon$ is an arbitrary small positive real number and $R$, $R_e$ are the rates of the message and equivocation. The capacity-equivocation region is defined as the convex closure of all achievable rate pairs $(R, R_e)$.

*Definition 3.* The secrecy capacity is defined as the maximum rate at which the confidential message can be sent to the

receiver in perfect secrecy; that is, $H(M) = H(M \mid Z^N)$. The secrecy capacity is

$$C_s = \max_{(R,R_e=R)\in\mathscr{R}} R, \tag{3}$$

where $\mathscr{R}$ is the capacity-equivocation region.

The capacity-equivocation regions of the model in Figure 1 with causal and noncausal channel states are shown in Theorems 4 and 5, respectively; see Section 2.3.

*2.3. Main Results*

**Theorem 4.** *For the wiretap channel with causal action-dependent states and rate-limited feedback, the capacity-equivocation region is the set*

$$\begin{aligned}\mathscr{R}_c = \{ (R, R_e) : 0 \le R_e \le R; R \le I(U;Y); R_e \le I(U;Y) \\ -I(U;Z) + R_f; R_e \le H(A \mid Z)\},\end{aligned} \tag{4}$$

*where $(A,U) \to (X,S) \to Y \to Z$ form a Markov chain and $R_f$ is the rate of the feedback link.*

*Comments.* (i) The proof of Theorem 4 is given in Appendices A and B. In Appendix A, a coding scheme is provided to show the achievability of the rate pair in $\mathscr{R}_c$. The proof of the converse part is shown in Appendix B. (ii) The set $\mathscr{R}_c$ is convex, and the proof is similar to that of [8, lemma 5]. (iii) To exhaust $\mathscr{R}_c$, it is enough to restrict $\mathscr{U}$ to satisfy $\|\mathscr{U}\| \le \|\mathscr{A}\|\|\mathscr{X}\|\|\mathscr{S}\| + 1$. It can be easily proved by using the *support lemma* [21, page 310].

**Theorem 5.** *For the wiretap channel with noncausal action-dependent states and rate-limited feedback, the capacity-equivocation region is the set*

$$\begin{aligned}\mathscr{R}_n = \{(R, R_e) : 0 \le R_e \le R; R \le I(U;Y) - I(U;S \mid A); \\ R_e \le I(U;Y) - I(U;Z) + R_f; \\ R_e \le H(A \mid Z)\},\end{aligned} \tag{5}$$

*where $(A,U) \to (X,S) \to Y \to Z$ form a Markov chain and $R_f$ is the rate of the feedback link.*

*Comments.* (i) The proof of Theorem 5 is given in Appendices C and D. In Appendix C, a coding scheme is provided to show the achievability of the rate pair in $\mathscr{R}_n$. The proof of the converse part is shown in Appendix D. (ii) The set $\mathscr{R}_n$ is convex, and the proof is similar to that of [8, lemma 5]. (iii) To exhaust $\mathscr{R}_n$, it is enough to restrict $\mathscr{U}$ to satisfy $\|\mathscr{U}\| \le \|\mathscr{A}\|\|\mathscr{X}\|\|\mathscr{S}\| + 2$. It can be easily proved by using the *support lemma* [21, page 310].

## 3. Discussion and Simulation

In this section, we first calculate the secrecy capacity of our model and show how the secrecy capacities of several existing channel models are derived from our results. Then, to better illustrate how our approach improves the existing results, we consider a binary symmetric channel with causal action-dependent states and rate-limited feedback. We try to compare the secrecy capacities of our model and the existing channel models through simulation.

*3.1. Discussion.* According to the definition in (3), the secrecy capacity of wiretap channel with causal action-dependent states and rate-limited feedback is

$$\begin{aligned}C_{\mathrm{sc}}(R_f) &= \max_{(R,R_e=R)\in\mathscr{R}_c} R \\ &= \max\min\{I(U;Y), I(U;Y) - I(U;Z) \\ &\qquad + R_f, H(A \mid Z)\}.\end{aligned} \tag{6}$$

We see that the secrecy capacity is a function of the feedback rate $R_f$. By setting $R_f = H(Y \mid UZ)$,

$$\begin{aligned}&I(U;Y) - I(U;Z) + R_f \\ &\quad = I(U;Y) - I(U;Z) + H(Y \mid UZ) \\ &\quad = H(U \mid Z) - H(U \mid Y) + H(Y \mid UZ) \\ &\quad = H(U \mid Z) - H(U \mid YZ) + H(Y \mid UZ),\end{aligned} \tag{7}$$

$$\begin{aligned}&I(U;Y) - I(U;Z) + R_f \\ &\quad = I(U;Y \mid Z) + H(Y \mid UZ) \\ &\quad = H(Y \mid Z) - H(Y \mid UZ) + H(Y \mid UZ) \\ &\quad = H(Y \mid Z),\end{aligned} \tag{8}$$

where (7) is from the Markov chain $U \to Y \to Z$. Substituting (8) into (6), we have

$$\begin{aligned}C_{\mathrm{sc}} &= \max_{(R,R_e=R)\in\mathscr{R}_c} R \\ &= \max\min\{I(U;Y), H(Y \mid Z), H(A \mid Z)\},\end{aligned} \tag{9}$$

which is the secrecy capacity of the causal case in [18].

Similarly, according to the definition in (3), the secrecy capacity of wiretap channel with noncausal action-dependent states and rate-limited feedback is

$$\begin{aligned}C_{\mathrm{sn}}(R_f) &= \max_{(R,R_e=R)\in\mathscr{R}_n} R \\ &= \max\min\{I(U;Y) - I(U;S \mid A), I(U;Y) \\ &\qquad - I(U;Z) + R_f, H(A \mid Z)\}.\end{aligned} \tag{10}$$

By setting $R_f = H(Y \mid UZ)$, (10) turns into

$$\begin{aligned}C_{\mathrm{sn}} = \max\min\{&I(U;Y) - I(U;S \mid A), \\ &H(Y \mid Z), H(A \mid Z)\},\end{aligned} \tag{11}$$

which is the secrecy capacity of the noncausal case in [18].

We should emphasize that the feedback in [18] comes from the output of the main channel and is used as a shared key between the transmitter and receiver. This kind of feedback mechanism gives rise to potential danger of revealing the key to the wiretapper since the wiretapper can guess the output of the main channel indirectly to get the key. Our approach avoids this potential danger of information leakage by constructing a feedback link and generating the feedback independently of the main channel's output.

In addition, without considering the causal or noncausal action-dependent states in the model of Figure 1, the secrecy capacity turns into $C'_{sc}(R_f) = \max\min\{I(U;Y), I(U;Y) - I(U;Z) + R_f\}$ which coincides with the secrecy capacity of the (degraded) wiretap channel with rate-limited feedback [16].

To better illustrate how our approach improves these models, such as channel without feedback, with noiseless feedback [18], and without actions [16], we present the simulation in the following subsection.

*3.2. Simulation and Comparison.* In the example, we consider a binary symmetric channel with causal action-dependent states and rate-limited feedback. The channel model is shown in Figure 2. Let the main channel be a binary symmetric channel (BSC), and let its crossover probability be affected by the channel states. The wiretap channel is also assumed to be a BSC with crossover probability $q$. In a more accurate way, define

$$p(y \mid x, s = i) = \begin{cases} (1-p)(1-i) + pi, & \text{if } y = x, \\ (1-p)i + p(1-i), & \text{otherwise}, \end{cases}$$
$$p(z \mid y) = \begin{cases} 1-q, & \text{if } z = y, \\ q, & \text{otherwise}, \end{cases} \tag{12}$$

where $i \in \{0,1\}$, $0 \le p \le 1$, and $0 \le q \le 1$.

To simplify the math, let the channel from the action to the channel states be a BSC with crossover probability equal to 1.0; that is, the channel states are totally determined by the action sequence. Similar to the arguments in [1, 13, 18], the maximum values of $I(U;Y)$, $H(A \mid Z)$, and $I(U;Y) - I(U;Z)$ are achieved when $g : \mathcal{U} \to \mathcal{A}$ and $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ are deterministic mappings. This implies that $H(A \mid Z) = H(U \mid Z)$. We choose $g$ and $f$ as

$$g(u = i) = i,$$
$$f(u = i, s = j) = i + j \pmod 2, \tag{13}$$

where $i, j \in \{0,1\}$. Let $U \sim \text{Bernoulli}(\alpha)$, where $0 \le \alpha \le 1$. Then, the joint distribution $p(a,s,u,x,y,z) = p(z \mid y)p(y \mid x,s)p(x \mid u,s)p(s \mid a)p(a \mid u)p(u)$ can be calculated. Since

$$p(u,y) = \sum_{a,s,x,z} p(a,s,u,x,y,z),$$
$$p(u,z) = \sum_{a,s,x,y} p(a,s,u,x,y,z), \tag{14}$$

by taking some mathematical calculation, we can get

$$\begin{aligned} C_{sc}(R_f) &= \max_{p(u)} \min\{I(U;Y), I(U;Y) - I(U;Z) \\ &\qquad + R_f, H(A \mid Z)\} \\ &= \min\{1 - h(p), h(p*q) - h(p) + R_f, h(p*q)\}, \end{aligned} \tag{15}$$

where $p * q = p + q - 2pq$ and $h(p)$ is the binary entropy function; that is, $h(p) = -p\log p - (1-p)\log(1-p)$. The calculation of the above $C_{sc}(R_f)$ is based on the *information theoretic methods* which involve the knowledge of *Information Theory* and *Probability Theory*. They are standard techniques, so it is not difficult to obtain the result of formula (15). The computation complexity of the calculation process is low. Similar arguments for calculating secrecy capacity can be seen in [1, 18]. Figure 3 shows that $C_{sc}$ starts from $h(p*q) - h(p)$ and increases linearly with $R_f$ until it gets saturated at $\min\{1 - h(p), h(p*q)\}$ for feedback rate $R_f \ge \min\{1 - h(p*q), h(p)\}$.

When $R_f \ge \min\{1 - h(p*q), h(p)\}$ and $p$ is fixed, the value of $C_{sc}$ changing with $q$ is shown in Figure 4. It can be seen from Figure 4 that when the crossover probability of the main channel is $p = 0.5$, the secrecy capacity is equal to 0. This result is straightforward since the legitimate receiver cannot correctly decode the message when the main channel is in the "poorest" condition. When the main channel condition gets better, that is, $p$ increases from 0.5 to 1.0, the maximum value of $C_{sc}$ increases. The secrecy capacity reaches $1 - h(q)$ when the crossover probability of the main channel $p = 1.0$. This is because when the main channel is noiseless, the input of main channel is the same as receiver's observation. This implies that the input of the main channel can be seen as the input of the wiretap channel.

When no feedback is imposed, that is, $R_f = 0$, Figure 5 shows the secrecy capacity $C_{sc}$ changing with $p$ when $q$ is fixed. As can be seen from Figure 5, when the quality of the wiretap channel gets better, that is, $q$ increases from 0.5 to 1.0, the secrecy capacity decreases. This warns the transmitter to pay attention to the trade-off between the transmission rate and confidentiality. When the wiretap channel is noiseless ($q = 1.0$), the secrecy capacity $C_{sc}$ is zero no matter how good the quality of the main channel is. It results from the fact that the wiretapper has the same observation as the receiver. Note that when the crossover probability of the wiretap channel is $q = 0.5$, the secrecy capacity is $C_{sc} = 1 - h(p)$ which is the capacity of the main channel. This is because no information leakage emerges when the wiretap channel is in the "poorest" condition.

The comparison among our approach, the model without feedback, the model with dependent noiseless feedback [18], and the model without taking actions [16] is presented as follows.

*3.2.1. Feedback versus No Feedback.* The comparison on the secrecy capacity between the channel with feedback and the channel without feedback is shown in Figures 6 and 7. It

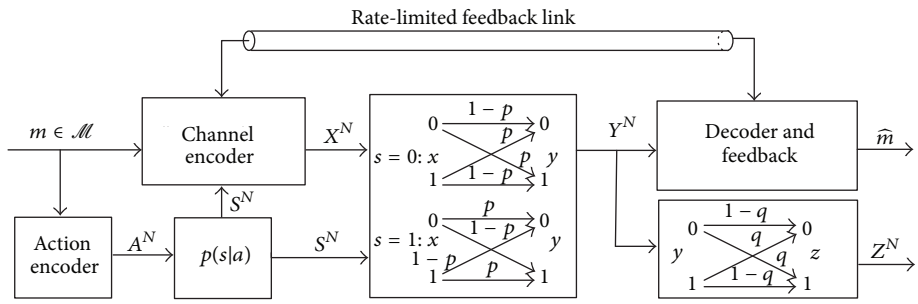


Figure 2: Binary symmetric channel with causal action-dependent states and rate-limited feedback.

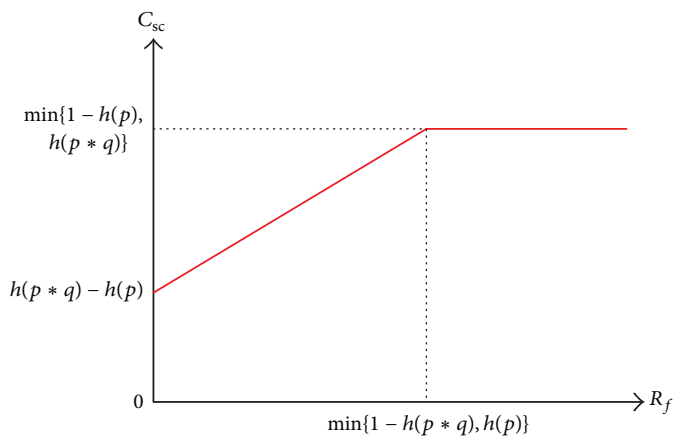


Figure 3: The secrecy capacity of the degraded binary symmetric wiretap channel with causal action-dependent states and rate-limited feedback.

can be seen from Figure 6 that the secrecy capacity of the model without feedback is covered by that with feedback. This indicates that the secrecy capacity of our approach is bigger than the channel model without feedback, and feedback can increase the secrecy capacity. To see it more clearly, we pick out four cross sections shown in Figure 7. In the lower right subgraph of Figure 7 where the wiretap channel is noiseless (i.e., $q = 1.0$), the secrecy capacity for the channel with and without feedback is a positive value and zero, respectively. This implies the fact that when the wiretap channel is noiseless and no feedback is imposed, no information can be securely transmitted to the receiver. Luckily, by introducing feedback, our approach makes it possible to transmit the message securely even if the wiretap channel is in the "best" condition.

*3.2.2. Independent Feedback versus Dependent Feedback.* Figures 8 and 9 present the secrecy capacities of our approach (with independent feedback) and the model [18] (with dependent feedback). In Figure 8, we can see that the secrecy capacity of our approach covers the secrecy capacity of [18]. To see it more clearly, we also choose four cross-sections shown in Figure 9. In general, the secrecy capacity of our approach is bigger than the model with noiseless feedback [18]. Concretely, one has the following.

(i) According to the results shown in [18], the secrecy capacity of the binary channel with causal channel states and noiseless feedback is $C_{\text{dai}} = \min\{1 - h(p), \min\{h(p * q), h(q)\}\}$. It is easy to see that $C_{\text{dai}} \leq C_{\text{sc}}(R_f)$ when $R_f \geq \min\{1 - h(p * q), h(p)\}$. The difference between $C_{\text{dai}}$ and $C_{\text{sc}}(R_f)$ is that there exists $h(q)$ in the expression of $C_{\text{dai}}$. Note that $h(q)$ is brought by $H(Y \mid Z)$ which is the wiretapper's uncertainty of the output of the main channel.

(ii) As can be seen from Figure 9, the maximum secrecy capacity gap between our approach and [18] is enlarged with the increase of $q$, and so does the range of variation of the main channel's transition probability when $C_{\text{sc}}(R_f) \geq C_{\text{dai}}$. This indicates that when the quality of the wiretap channel becomes better (i.e., $q$ increases from 0.5 to 1.0), the number of the main channels, in which our approach brings about bigger secrecy capacities than those in [18], increases.

(iii) In Figures 8 and 9, we also see that when $q$ increases (from 0.5 to 1.0), the maximum value of secrecy capacity in our approach and the model with noiseless feedback [18] decreases. This results from the fact that when $q$ varies from 0.5 to 1.0, the condition of the wiretap channel gets better so that the wiretapper is more able to obtain the confidential message.

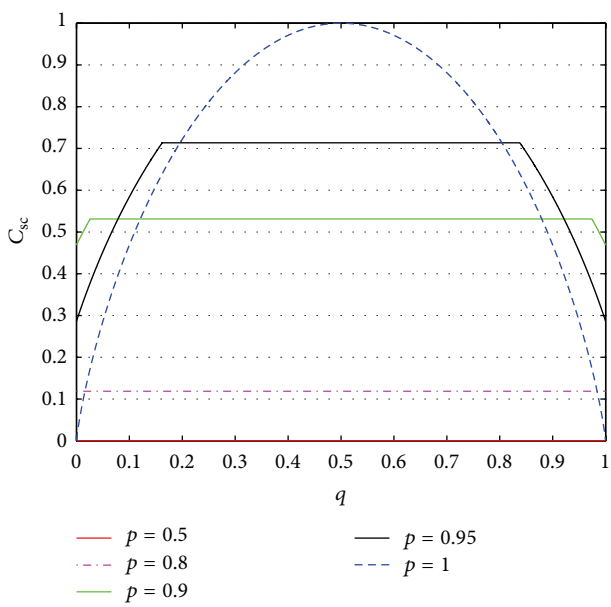


Figure 4: The secrecy capacity of the BSC example with feedback rate $R_f \geq \min\{1 - h(p * q), h(p)\}$.

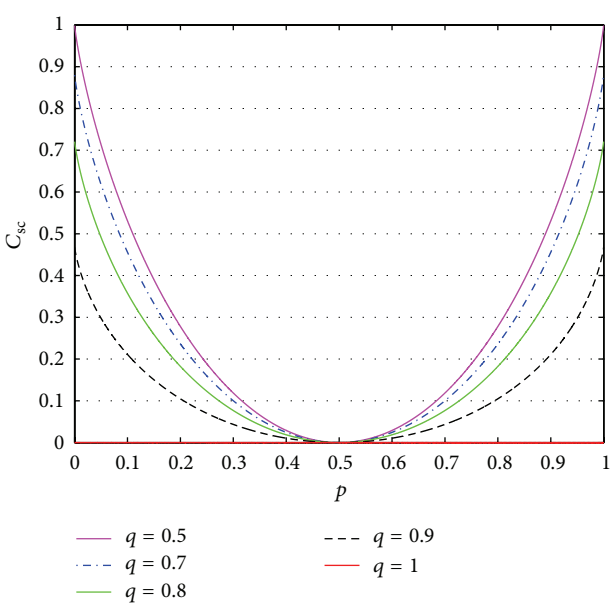


Figure 5: The secrecy capacity of the BSC example with feedback rate $R_f = 0$.

(iv) In the lower right subgraph of Figure 9 where $q = 1$ (i.e., the wiretap channel is noiseless), the secrecy capacity of the model with noiseless feedback [18] is zero. This indicates that it is unable to transmit the message securely over the channel. However, the secrecy capacity of our approach is positive. This means that the transmitter can still send the message securely although the wiretapper is more "powerful" than the legitimate receiver.

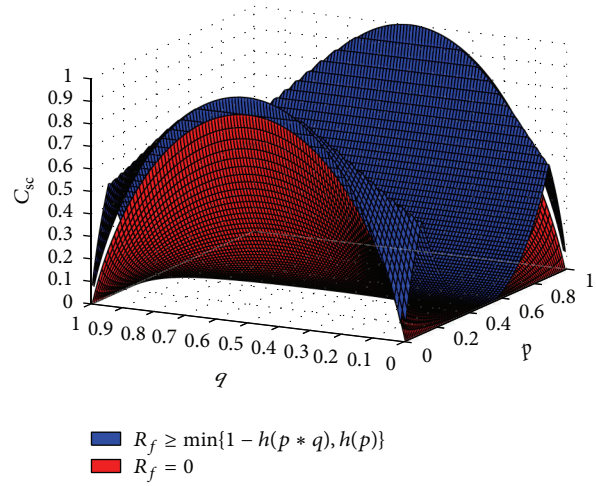


Figure 6: With feedback versus without feedback (i).

We can also explain the results from the aspect of feedback. In the channel model with noiseless feedback [18], the feedback is dependent on the output of the main channel and used as a shared key between the transmitter and the receiver. This enables the wiretapper to get the key and, further, the confidential message by guessing the output of the main channel indirectly. However, in our model, since the feedback is independent of the channel output, the wiretapper tries in vain to get the key through guessing the channel output. Therefore, our approach makes the system more secure.

*3.2.3. Actions versus No Actions.* We consider six representative cases to evaluate the secrecy capacities of our model and the model without actions [16].

*Case 1* ($p = 0.99, q = 0.65$). The main channel is almost "perfect," while the wiretap channel is in the "poorest" condition.

*Case 2* ($p = 0.96, q = 0.82$). The main channel is better than the wiretap channel. They are both in "good" condition.

*Case 3* ($p = 0.93, q = 0.93$). The main channel and wiretap channel share the same transition probability. They are both in "good" condition.

*Case 4* ($p = 0.90, q = 0.99$). The wiretap channel is better than the main channel. They are both in "good" condition.

*Case 5* ($p = 0.65, q = 0.99$). The wiretap channel is almost "perfect," while the main channel is in the "poorest" condition.

*Case 6* ($p = 0.60, q = 0.60$). The main channel and wiretap channel share the same transition probability. They are both in the "poorest" condition.

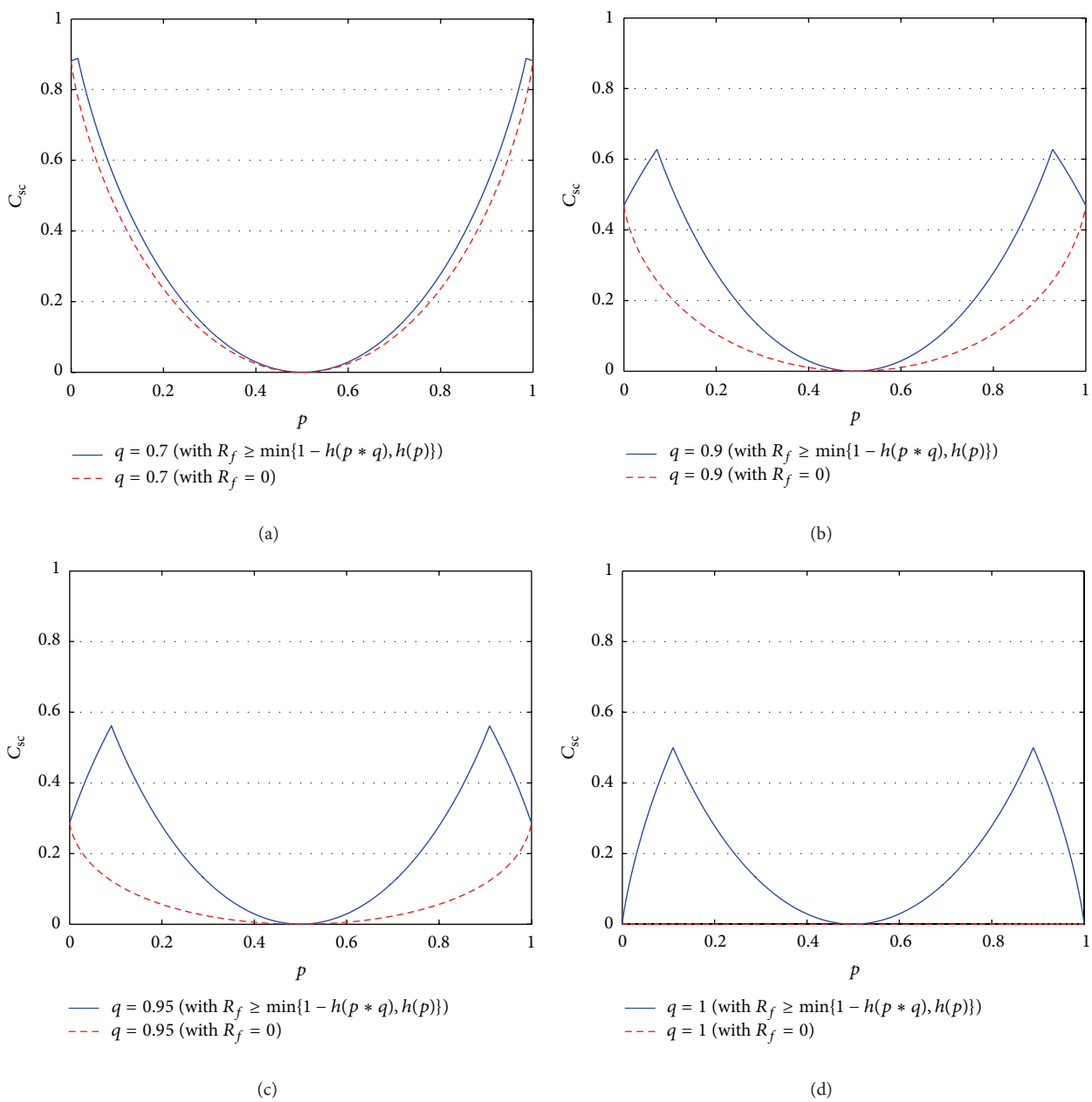


Figure 7: With feedback versus without feedback (ii).

Our model provides insights into the value of actions in the two-stage coding systems, such as recording for magnetic storage devices and coding for computer memories with defects. To better explain the value of actions, let us consider an example of writing information into a memory with defects. The memory can be seen as the main channel in our model, the location of the defects corresponds to the channel states, and writing information into the memory corresponds to transmitting information over the channel. Before writing information into the memory, suppose that neither encoder nor decoder knows the locations of the defects. In the first stage, the encoder writes into the memory for the first time. It gets a noisy version of the inputs when it reads from the memory. Then, in the second stage, the encoder rewrites at whichever memory locations it chooses before the decoder attempts to decode the information [1]. Note that, in the second stage, the encoder will have some knowledge about the memory defects according to the information written in the first stage and the noisy version it reads from the memory in that stage. The first stage can be seen as the "preparing stage" in which the encoder takes "actions" to learn about the defects (i.e., channel state).

Figure 10 shows the process of writing eight data blocks into a memory with defects by our approach and the approach in [16] where no actions were taken. As can be seen from Figure 10, without taking actions to probe the locations of the

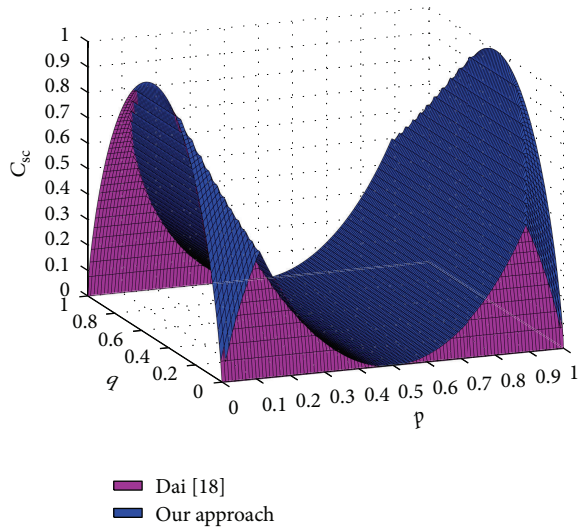


Figure 8: Independent feedback versus dependent feedback (i).

Table 1: The extra overhead of the transmission time.

| Case | The extra overhead of transmission time |
|---|---|
| 1 | 0 |
| 2 | 3.37% |
| 3 | 13.64% |
| 4 | 7.51% |
| 5 | 0 |
| 6 | 0 |

defects, some data blocks are missed after they are written into the memory. It is well known that data integrity is one of the most important aspects of communication. From the above example, by introducing actions in such two-stage systems, the integrity of the transmitted data is guaranteed. Therefore, the scheme in [16] where no actions were taken is unsuitable for the two-stage systems. From this point of view, our assumption that the encoder takes actions to acquire channel state information is essential and valuable.

At the same time, the cost of taking actions is that the secrecy capacity decreases; see Figure 11. The secrecy capacity of the binary example without actions [16] is $C'_{sc}(R_f) = \min\{1 - h(p), h(p * q) - h(p) + R_f\}$. As can be seen from Figure 11, when the rate of the feedback link $R_f$ is less than a specific value, the secrecy capacity $C_{sc}$ of our approach is equal to the secrecy capacity $C'_{sc}$ of the model [16]. This means no extra transmission time (or channel use) is produced by introducing "actions" when $R_f$ is small. However, when $R_f$ is greater than a specific value, the secrecy capacity is $C_{sc} \leq C'_{sc}$. This indicates that the "actions" bring the overhead of transmission time. Concretely, the extra overhead of the transmission time is shown in Table 1.

From Table 1, we can see that when $p = 0.96$, $q = 0.82$ (Case 2), the overhead of the transmission time shows an increase of 3.37 percent over the approach [16] where no action-dependent states were considered. When $p = 0.93$, $q = 0.93$ and $p = 0.90$, $q = 0.99$, the extra time overhead is 13.64 and 7.51 percent, respectively. In general, the extra time overhead is small. On the premise of data integrity, the amount of such extra time overhead is tolerable.

## 4. Conclusion

This paper studies the wiretap channel with action-dependent states and rate-limited feedback. It is a degraded channel model where the wiretap channel is degraded from the main channel. The capacity-equivocation region of such a channel both for the case where the channel inputs depend noncausally on the state sequence and the case where they are restricted to causal dependence is obtained. This paper also gets the secrecy capacities for both cases. At the same time, we construct capacity-achieving coding schemes using the methods of Wyner's random binning, Gel'fand and Pinsker's coding technique, and rate splitting. The simulation results show that using feedback can increase the secrecy capacity of the channel in Figure 1, and the secrecy capacity of our model is bigger than that of [16, 18]. Besides, we also find that taking actions to affect the channel states can ensure the data integrity of the message transmitted in the two-stage systems although the tolerable overhead of transmission time is brought.

Some potential directions that our work leaves open for future study are as follows.

(i) *Nondegraded Wiretap Channel*. In this paper, the wiretap channel is degraded from the main channel where $p(z, y \mid x, s) = p(z \mid y)p(y \mid x, s)$. This indicates that $(X, S) \to Y \to Z$ form a Markov chain. However, the degraded wiretap channel is a special case of non-degraded wiretap channel where $p(z, y \mid x, s) = p(z \mid y)p(y \mid x, s)$ does not need to hold. Without this Markov chain, the coding schemes as well as the proof of converse part will be different, and further the secrecy capacity of the non-degraded wiretap channel will be different from the current results.

(ii) *Adaptive Action*. Adaptive action means that the action sequence is generated by the message and the previous channel states, that is, $a_i(m, s^{i-1})$. Adaptive action is widely used in many applications such as information hiding, digital watermarking, and data storage in the memory. It is valuable to study the adaptive action in our model. From [22], we have already known that adaptive action is not useful in increasing the point-to-point channel capacity. We will study whether it influences the secrecy capacity of our channel model.

(iii) *Multiple Transmitters or Receivers*. Nowadays, many types of communication involves two or more transmitters (or receivers), such as the multiple-access channel (MAC), broadcast channel (BC), and multiple-input-multiple-output (MIMO) channel. By introducing action-dependent states and feedback in such channels (with an additional wiretapper), we

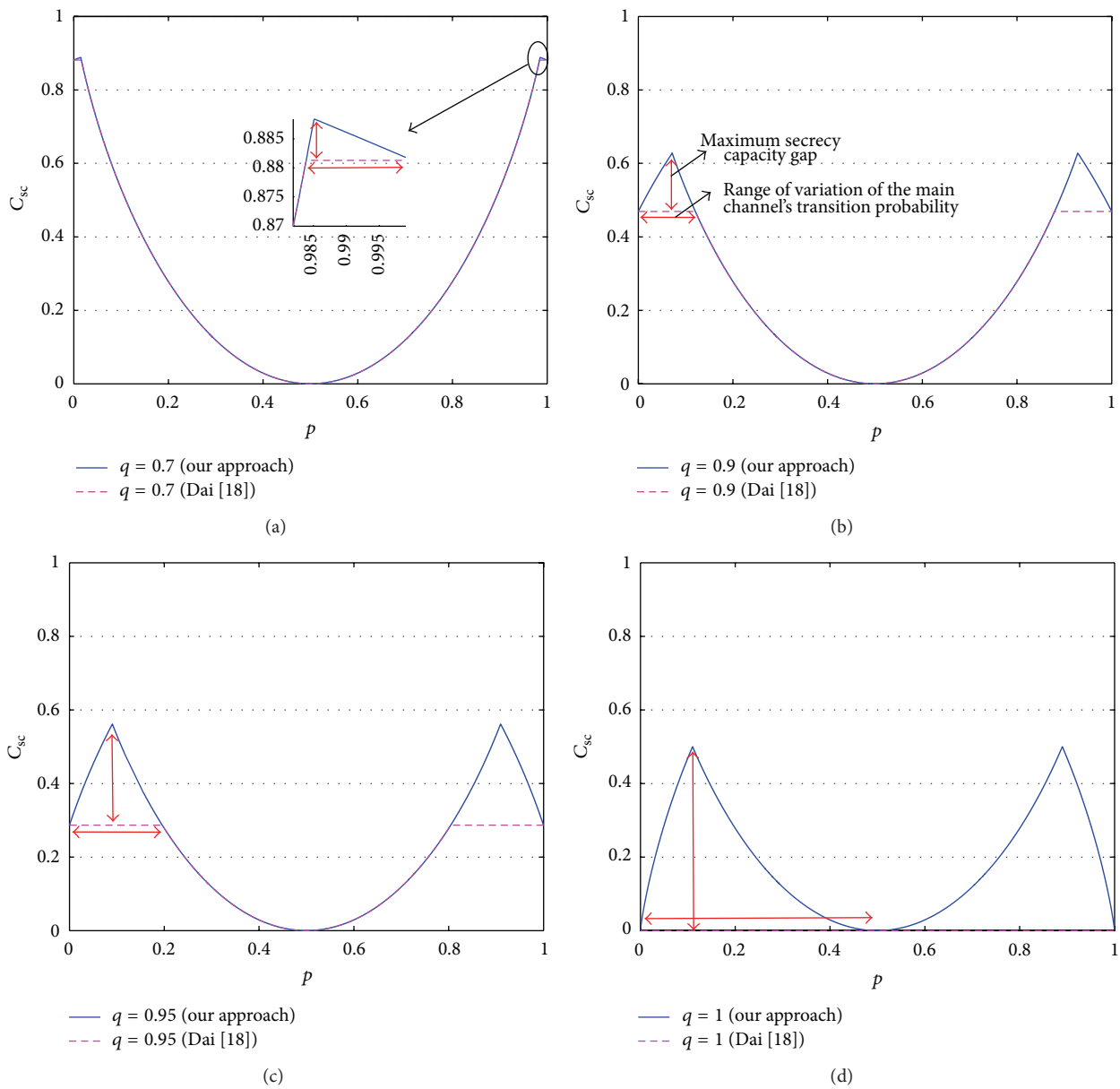


Figure 9: Independent feedback versus dependent feedback (ii).


Figure 10: Writing information into a memory with defects.

can revisit their capacity-equivocation regions and secrecy capacities.

Besides, the Gaussian wiretap channel with action-dependent states and feedback is also a practical channel model that is worthy of being explored.

## Appendices

### A. Coding Schemes for Causal Channel State Information

This section provides the coding schemes to achieve $(R, R_e) \in \mathscr{R}_c$. Similar to the coding scheme in [18], two different coding schemes for $H(A \mid Z) \geq \min\{I(U;Y), I(U;Y) - I(U;Z) + R_f\}$ and $H(A \mid Z) \leq \min\{I(U;Y), I(U;Y) - I(U;Z) + R_f\}$ are constructed in Cases 1 and 2, respectively.

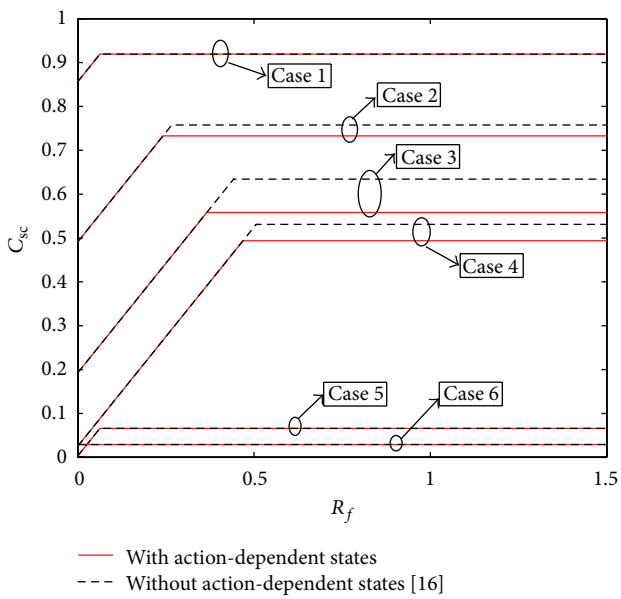


FIGURE 11: Actions versus no actions.

*Case 1* ($H(A \mid Z) \geq \min\{I(U;Y), I(U;Y) - I(U;Z) + R_f\}$). In Case 1, we need to show that all rate pairs $(R, R_e)$, satisfying

$$\begin{aligned} 0 &\leq R_e \leq R, \\ R &\leq I(U;Y), \\ R_e &\leq I(U;Y) - I(U;Z) + R_f, \end{aligned} \tag{A.1}$$

are achievable. It is sufficient to prove that the rate pairs $(R, R_e = \min\{I(U;Y), I(U;Y) - I(U;Z) + R_f\})$ are achievable. In the coding scheme, Wyner's random binning technique and rate splitting are used. The feedback symbol is used as a shared secret key between the transmitter and the receiver. The wiretapper is assumed to have no knowledge of the secret key.

Split the message $\mathscr{M}$ into two parts, that is, $\mathscr{M} = (\mathscr{M}_1, \mathscr{M}_2)$. The corresponding random variables $M_1$, $M_2$ are uniformly distributed over $\mathscr{M}_1 = \{1, 2, 3, \ldots, 2^{NR'}\}$ and $\mathscr{M}_2 = \{1, 2, 3, \ldots, 2^{NR'_f}\}$, where

$$\begin{aligned} 0 &\leq R'_f \leq \min\left\{I(U;Z), R_f\right\}, \\ R' &= R - R'_f > 0. \end{aligned} \tag{A.2}$$

Define three alphabets $\mathscr{J}_N$, $\mathscr{L}_N$, and $\mathscr{F}_N$ satisfying

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \log \|\mathscr{J}_N\| &= I(U;Z) - R'_f; \\ \lim_{N\to\infty} \frac{1}{N} \log \|\mathscr{L}_N\| &= I(U;Y) - I(U;Z); \\ \lim_{N\to\infty} \frac{1}{N} \log \|\mathscr{F}_N\| &= R'_f. \end{aligned} \tag{A.3}$$

Note that $\lim_{N\to\infty}((1/N)\log(\|\mathscr{J}_N\| \cdot \|\mathscr{L}_N\| \cdot \|\mathscr{F}_N\|)) = I(U;Y)$.

Since $R \leq I(U;Y)$, let $\mathscr{M}_1 = \mathscr{D}_N \times \mathscr{L}_N$ and $\mathscr{M}_2 = \mathscr{F}_N$, where $\mathscr{D}_N$ is an arbitrary set such that $\lim_{N\to\infty}((1/N)\log\|\mathscr{M}\|) = R$ holds and $\|\mathscr{D}_N\| \leq \|\mathscr{J}_N\|$. Define a deterministic mapping $g_1$ from $\mathscr{J}_N$ into $\mathscr{D}_N$, which partitions $\mathscr{J}_N$ into subsets of size $\|\mathscr{J}_N\|/\|\mathscr{D}_N\|$. The codebook is constructed as follows.

*Codebook Generation and Encoding.* Generate $2^{NR'}$ independent and identically distributed (i.i.d) action sequences $a^N(m_1)$ according to $p(a^N) = \prod_{i=1}^N p(a_i)$, where $m_1 \in \mathscr{M}_1$. The channel state sequence $s^N$ is generated through the discrete memoryless channel $p(s^N \mid a^N) = \prod_{i=1}^N p(s_i \mid a_i)$.

The feedback link is shared between the transmitter and the legitimate receiver. It is assumed that a secret key $k \in \{1, 2, 3, \ldots, 2^{NR'_f}\}$ is transmitted through the feedback link. Let the corresponding random variable $K$ be independent of $M$ and uniformly distributed over $\{1, 2, 3, \ldots, 2^{NR'_f}\}$.

For each $a^N(m_1)$, generate $2^{N(I(U;Y)-R')}$ independent and identically distributed (i.i.d) sequences $u^N$ according to $p(u^N \mid a^N) = \prod_{i=1}^N p(u_i \mid a_i)$. Put these sequences into $2^{NR'_f}$ bins so that each bin contains $\|\mathscr{J}_N\|/\|\mathscr{D}_N\|$ sequences $u^N(m_1, t_b, t_u)$, where the bin index $t_b \in \{1, 2, 3, \ldots, 2^{NR'_f}\}$ and the sequence index $t_u \in \{1, 2, \ldots, \|\mathscr{J}_N\|/\|\mathscr{D}_N\|\}$.

Suppose that $k \in \{1, 2, 3, \ldots, 2^{NR'_f}\}$ is the shared secret key between the transmitter and the receiver during one transmission. To send message $(m_1, m_2) = (d, l, m_2)$ with the action sequence $a^N(m_1)$ and the state sequence $s^N$, where $d \in \mathscr{D}_N$, $l \in \mathscr{L}_N$, and $m_2 \in \mathscr{M}_2$, randomly choose a sequence $u^N(m_1, t_b, t_u)$ from the bin indexed by $t_b = m_2 \oplus k$. Here $\oplus$ is modulo addition over $\mathscr{F}_N$. Then, the input sequence of the main channel is generated by $p(x^N \mid u^N, s^N) = \prod_{i=1}^N p(x_i \mid u_i, s_i)$.

*Decoding.* When observing the channel output $y^N$, the receiver tries to find a sequence $u^N(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u)$ such that $(u^N(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u), y^N) \in T^N_{UY}$. Then, the decoder outputs $(\widehat{m}_1, \widehat{t}_b \ominus k)$, where $\ominus$ is modulo subtraction over $\mathscr{F}_N$. If no such $(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u)$ exists, take $(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u) = (1, 1, 1)$.

*Analysis of Error Probability.* By using similar arguments in [1], it is easy to show that the legitimate receiver can decode the message with vanishing probability of decoding error. The details for proving $P_e \leq \epsilon$ are omitted.

*Analysis of Equivocation.* We focus on calculating wiretapper's equivocation of the message, that is, the uncertainty about the message given wiretapper's observation. To serve the analysis of equivocation, the fact that $M_2$ is independent of $M_2 \oplus K$ will be shown firstly. Using similar derivation in [18], we start with the definition of mutual independence

$$\begin{aligned} &p\left(M_2 \oplus K = k', M_2 = m_2\right) \\ &\quad = p\left(K = k' \ominus m_2, M_2 = m_2\right) \\ &\quad =^{(a1)} p\left(K = k' \ominus m_2\right) p\left(M_2 = m_2\right) \end{aligned}$$

$$
\begin{aligned}
&=^{(a2)} \frac{1}{\|\mathscr{F}_N\|} \cdot \frac{1}{\|\mathscr{F}_N\|}, \\
&p\left(M_2 \oplus K = k'\right) \\
&\quad = \sum_k p\left(M_2 \oplus K = k' \mid K = k\right) p\left(K = k\right) \\
&\quad =^{(a3)} \sum_k p\left(M_2 \oplus K = k' \mid K = k\right) \frac{1}{\|\mathscr{F}_N\|} \\
&\quad = \frac{1}{\|\mathscr{F}_N\|} \sum_k p\left(M_2 \oplus K = k' \mid K = k\right) \\
&\quad = \frac{1}{\|\mathscr{F}_N\|} \sum_k p\left(M_2 = k' \ominus k \mid K = k\right) \\
&\quad =^{(a4)} \frac{1}{\|\mathscr{F}_N\|} \sum_k p\left(M_2 = k' \ominus k\right) \\
&\quad = \frac{1}{\|\mathscr{F}_N\|},
\end{aligned}
\tag{A.4}
$$

where $(a1)$ and $(a4)$ follow from the fact that $M_2$ is independent of $K$ and $(a2)$ and $(a3)$ follow from the fact that $M_2$ and $K$ are both uniformly distributed over $\mathscr{F}_N$. According to (A.4),

$$
\begin{aligned}
&p\left(M_2 \oplus K = k', M_2 = m_2\right) \\
&\quad = p\left(M_2 \oplus K = k'\right) p\left(M_2 = m_2\right).
\end{aligned}
\tag{A.5}
$$

Therefore, $M_2$ is independent of $M_2 \oplus K$.

Utilizing the above fact,

$$
\begin{aligned}
H\left(M \mid Z^N\right) &= H\left(M_1, M_2 \mid Z^N\right) \\
&= H\left(M_1 \mid Z^N\right) + H\left(M_2 \mid Z^N, M_1\right) \\
&\geq H\left(M_1 \mid Z^N\right) + H\left(M_2 \mid Z^N, M_1, K \oplus M_2\right) \\
&=^{(a5)} H\left(M_1 \mid Z^N\right) + H\left(M_2 \mid K \oplus M_2\right) \\
&=^{(a6)} H\left(M_1 \mid Z^N\right) + H\left(M_2\right) \\
&=^{(a7)} H\left(M_1 \mid Z^N\right) + NR'_f,
\end{aligned}
\tag{A.6}
$$

where $(a5)$ follows from the Markov chain $M_2 \to M_2 \oplus K \to (Z^N, M_1)$, $(a6)$ follows from the fact that $M_2$ is independent of $M_2 \oplus K$, and $(a7)$ follows from the fact that $M_2$ is uniformly distributed over $\{1, 2, 3, \ldots, 2^{NR'_f}\}$. To bound $H(M_1 \mid Z^N)$ in (A.6), Csiszar's method for analyzing equivocation [8] is used:

$$
\begin{aligned}
&H\left(M_1 \mid Z^N\right) \\
&\quad = H\left(M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad = H\left(M_1, Z^N, U^N\right) - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad = H\left(M_1, U^N\right) + H\left(Z^N \mid M_1, U^N\right) \\
&\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad \geq H\left(U^N\right) + H\left(Z^N \mid M_1, U^N\right) - H\left(U^N \mid M_1, Z^N\right) \\
&\qquad - H\left(Z^N\right) \\
&\quad \geq H\left(U^N\right) - H\left(U^N \mid Y^N\right) + H\left(Z^N \mid M_1, U^N\right) \\
&\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad = I\left(U^N; Y^N\right) + H\left(Z^N \mid M_1, U^N\right) \\
&\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad =^{(a8)} I\left(U^N; Y^N\right) + H\left(Z^N \mid U^N\right) \\
&\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad =^{(a9)} NI\left(U; Y\right) + NH\left(Z \mid U\right) \\
&\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\
&\quad \geq^{(a10)} NI\left(U; Y\right) + NH\left(Z \mid U\right) \\
&\qquad - H\left(U^N \mid M_1, Z^N\right) - NH\left(Z\right) \\
&\quad \geq^{(a11)} NI\left(U; Y\right) - NI\left(U; Z\right) - N\epsilon_2.
\end{aligned}
\tag{A.7}
$$

In the above derivations, $(a8)$ follows from the fact that $M_1 \to U^N \to Z^N$ is a Markov chain. $(a9)$ follows from the fact that $S^N$, $U^N$ and $X^N$ are i.i.d generated and both the main channel and wiretap channel are discrete memoryless. $(a10)$ follows from the fact that $H(Z^N) = \sum_{i=1}^{N} H(Z_i \mid Z^{i-1}) \leq NH(Z)$, where $Z$ is regarded as the general random variable of $Z_i$. To verify $(a11)$, note the fact that given the message $m_1 = (d, l)$, the number of $U^N$ is

$$
\begin{aligned}
2^{N(I(U;Y)-R')} &= \frac{2^{NI(U;Y)}}{2^{NR'}} \\
&= \frac{\|\mathscr{J}_N\| \cdot \|\mathscr{L}_N\| \cdot \|\mathscr{F}_N\|}{\|\mathscr{M}_1\|} \\
&= \frac{\|\mathscr{J}_N\| \cdot \|\mathscr{L}_N\| \cdot \|\mathscr{F}_N\|}{\|\mathscr{D}_N\| \cdot \|\mathscr{L}_N\|}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\|\mathscr{I}_N\| \cdot \|\mathscr{F}_N\|}{\|\mathscr{D}_N\|} \\
&\le \|\mathscr{I}_N\| \cdot \|\mathscr{F}_N\| \\
&= 2^{N(I(U;Z)-R'_f)} \cdot 2^{NR'_f} \\
&= 2^{NI(U;Z)}.
\end{aligned} \tag{A.8}
$$

By applying the standard channel coding theorem [23, Theorem 7.7.1], the wiretapper can decode $U^N$ with vanishing decoding error probability. Then, utilizing Fano's inequality, we get $H(U^N \mid M_1, Z^N) \le N\epsilon_2$, where $\epsilon_2 \to 0$ when $N \to \infty$. (A.7) is verified.

Using the definition of equivocation and substituting (A.7) into (A.6),

$$
\begin{aligned}
\lim_{N\to\infty} \Delta &= \lim_{N\to\infty} \frac{H\left(M \mid Z^N\right)}{N} \\
&\ge \lim_{N\to\infty} \left( \frac{H\left(M_1 \mid Z^N\right)}{N} + R'_f \right) \\
&\ge \lim_{N\to\infty} \left( \frac{NI(U;Y) - NI(U;Z) - N\epsilon_2}{N} + R'_f \right) \\
&= I(U;Y) - I(U;Z) + R'_f \\
&\ge R_e.
\end{aligned} \tag{A.9}
$$

This completes the achievability proof of Case 1.

*Case 2* ($H(A \mid Z) \le \min\{I(U;Y), I(U;Y) - I(U;Z) + R_f\}$). In Case 2, we show that all rate pairs $(R, R_e)$, satisfying

$$
\begin{aligned}
0 &\le R_e \le R, \\
R &\le I(U;Y), \\
R_e &\le H(A \mid Z),
\end{aligned} \tag{A.10}
$$

are achievable. It is sufficient to prove that the rate pairs $(R, R_e = \min\{I(U;Y), H(A \mid Z)\})$ are achievable. The coding scheme is similar to [18].

*Codebook Generation and Encoding.* Generate $2^{NR}$ action sequences $a^N(m)$ according to $p(a^N) = \prod_{i=1}^N p(a_i)$, where $m \in \{1, 2, \ldots, 2^{NR}\}$. The state $s^N$ is generated in response to $a^N(m)$ according to $p(s^N \mid a^N) = \prod_{i=1}^N p(s_i \mid a_i)$. Generate $2^{NR}$ i.i.d sequences $u^N(m)$ according to $p(u^N \mid a^N) = \prod_{i=1}^N p(u_i \mid a_i)$. To send message $m$ with the action sequence $a^N(m)$ and the state sequence $s^N$, find the sequence $u^N(m)$. Then, the input sequence of the main channel is generated by $p(x^N \mid u^N, s^N) = \prod_{i=1}^N p(x_i \mid u_i, s_i)$.

*Decoding.* When observing the channel output $y^N$, the receiver tries to find a sequence $u^N(\widehat{m})$ such that $(u^N(\widehat{m}), y^N) \in T_{UY}^N(\epsilon'_1)$. Then, the decoder outputs $\widehat{m}$. If no such $\widehat{m}$ exists, take $\widehat{m} = 1$.

*Analysis of Error Probability.* By using similar arguments in [1], it is easy to show that the legitimate receiver can decode the message with vanishing probability of decoding error. The details for proving $P_e \le \epsilon$ are omitted.

*Analysis of Equivocation.* We focus on calculating wiretapper's equivocation of the the message, that is, the uncertainty about the message given wiretapper's observation. The method for calculating the equivocation in [13] is utilized in this case:

$$
\begin{aligned}
\lim_{N\to\infty} \Delta &= \lim_{N\to\infty} \frac{H\left(M \mid Z^N\right)}{N} \\
&=^{(a12)} \lim_{N\to\infty} \frac{H\left(A^N \mid Z^N\right)}{N} \\
&=^{(a13)} \lim_{N\to\infty} \frac{NH(A \mid Z)}{N} \\
&= H(A \mid Z) \\
&\ge R_e,
\end{aligned} \tag{A.11}
$$

where $(a12)$ follows from the fact that the action sequence $a^N(m)$ is a determination function of the message $m$ and $(a13)$ follows from the fact that $A^N$ and $X^N$ are i.i.d generated and both the main channel and wiretap channel are discrete memoryless. This completes the achievability proof of Case 2.

The achievability proof of Theorem 4 is completed.

## B. Proof of the Converse Part of Theorem 4

In this section, we show that all achievable rate pairs $(R, R_e)$ for the channel model in Figure 1 with causal states are contained in $\mathscr{R}_c$. Concretely, for any rate pair $(R, R_e)$, feedback alphabet $\mathscr{K}$, and decoding error probability $P_e$ satisfying

$$R = \lim_{N\to\infty} \frac{\log \|\mathscr{M}\|}{N}, \tag{B.1}$$

$$\lim_{N\to\infty} \frac{\log \|\mathscr{K}\|}{N} \le R_f, \tag{B.2}$$

$$R_e \le \lim_{N\to\infty} \Delta, \tag{B.3}$$

$$P_e \le \epsilon, \tag{B.4}$$

there exist random variables $(A, U) \to (X, S) \to Y \to Z$ such that

$$0 \le R_e \le R, \tag{B.5}$$

$$R \le I(U;Y), \tag{B.6}$$

$$R_e \le I(U;Y) - I(U;Z) + R_f, \tag{B.7}$$

$$R_e \le H(A \mid Z). \tag{B.8}$$

The four inequalities (B.5), (B.6), (B.7), and (B.8) will be proved successively.

One has consider condition (B.5) first.

$$\begin{aligned} R_e &\le \lim_{N\to\infty} \Delta \\ &= \lim_{N\to\infty} \frac{H\left(M \mid Z^N\right)}{N} \\ &\le \lim_{N\to\infty} \frac{H(M)}{N} \\ &=^{(b1)} R, \end{aligned} \tag{B.9}$$

where $(b1)$ is from (B.1). Condition (B.5) is proved.

To prove condition (B.6), we consider

$$\begin{aligned} \frac{1}{N}H(M) &= \frac{1}{N}I\left(M;Y^N\right) + \frac{1}{N}H\left(M \mid Y^N\right) \\ &\le^{(b2)} \frac{1}{N}I\left(M;Y^N\right) + \frac{1}{N}\eta\left(P_e\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} I\left(M;Y_i \mid Y^{i-1}\right) + \frac{1}{N}\eta\left(P_e\right) \\ &\le \frac{1}{N}\sum_{i=1}^{N} I\left(M,Y^{i-1},S^{i-1};Y_i\right) + \frac{1}{N}\eta\left(P_e\right) \\ &\le^{(b3)} \frac{1}{N}\sum_{i=1}^{N} I\left(U_i;Y_i\right) + \frac{1}{N}\eta\left(P_e\right). \end{aligned} \tag{B.10}$$

Inequality $(b2)$ is from Fano's inequality. To be more accurate, $\eta(P_e) = h(P_e) + P_e \log(\|M\| - 1)$, where $h(P_e)$ is the binary entropy function; that is, $h(P_e) = -P_e \log P_e - (1 - P_e)\log(1 - P_e)$. Note that $\eta(P_e) \to 0$ when $P_e$ approaches zero. $(b3)$ is from defining $U_i = (M, Y^{i-1}, S^{i-1})$. According to (B.1) and (B.10), it is easy to see that

$$\begin{aligned} R &= \lim_{N\to\infty} \frac{\log\|\mathscr{M}\|}{N} \\ &=^{(b4)} \lim_{N\to\infty} \frac{1}{N}H(M) \\ &\le \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} I\left(U_i;Y_i\right) + \frac{1}{N}\eta\left(P_e\right), \end{aligned} \tag{B.11}$$

where $(b4)$ follows from the fact that $M$ is uniformly distributed over $\{1, 2, \ldots, \|\mathscr{M}\|\}$.

To prove condition (B.7), the recursive argument lemma [16] will be exploited. Therefore, we present the lemma below. The detailed proof of the lemma is shown in [16].

*Recursive Argument Lemma [16].* For each $j = 1, 2, \ldots, N$, the following inequality holds:

$$\begin{aligned} &H\left(K^j \mid Z^j\right) + I\left(M, X^j; Y^j \mid K^j, Z^j\right) \\ &\quad \le H\left(K^{j-1} \mid Z^{j-1}\right) + I\left(M, X^{j-1}; Y^{j-1} \mid K^{j-1}, Z^{j-1}\right) \\ &\qquad + I\left(X_j;Y_j \mid Z_j\right) + H\left(K_j \mid M, X^{j-1}, K^{j-1}, Z^{j-1}\right), \end{aligned} \tag{B.12}$$

where $K$, $X$, $Y$, and $Z$ are the the feedback random variable, the input of the main channel, the input of the wiretap channel, and the output of the wiretap channel. (Actually, [16] gave the lemma under a more relaxed condition that the wiretap channel did not need to be physically degraded.)

Then, we prove condition (B.7) as follows:

$$\begin{aligned} R_e &\le \lim_{N\to\infty} \Delta \\ &= \lim_{N\to\infty} \frac{1}{N}H\left(M \mid Z^N\right) \\ &= \lim_{N\to\infty} \frac{1}{N}\left(I\left(M;Y^N,K^N \mid Z^N\right)\right. \\ &\qquad \left. + H\left(M \mid Y^N, Z^N, K^N\right)\right) \\ &\le \lim_{N\to\infty} \frac{1}{N}\left(I\left(M;Y^N,K^N \mid Z^N\right) + H\left(M \mid Y^N\right)\right) \\ &\le^{(b5)} \lim_{N\to\infty} \frac{1}{N}\left(I\left(M;Y^N,K^N \mid Z^N\right) + \eta\left(P_e\right)\right) \\ &= \lim_{N\to\infty} \frac{1}{N}\left(I\left(M;K^N \mid Z^N\right)\right. \\ &\qquad \left. + I\left(M;Y^N \mid K^N, Z^N\right) + \eta\left(P_e\right)\right) \\ &\le \lim_{N\to\infty} \frac{1}{N}\left(H\left(K^N \mid Z^N\right)\right. \\ &\qquad \left. + I\left(M,U^N;Y^N \mid K^N, Z^N\right) + \eta\left(P_e\right)\right), \end{aligned} \tag{B.13}$$

where $(b5)$ follows from Fano's inequality. By applying the lemma recursively, the two terms in (B.13) can be single letterized as follows:

$$\begin{aligned} &H\left(K^N \mid Z^N\right) + I\left(M,U^N;Y^N \mid K^N, Z^N\right) \\ &\quad \le H\left(K^{N-1} \mid Z^{N-1}\right) + I\left(M,U^{N-1};Y^{N-1} \mid K^{N-1}, Z^{N-1}\right) \\ &\qquad + I\left(U_N;Y_N \mid Z_N\right) + H\left(K_N \mid M, U^{N-1}, K^{N-1}, Z^{N-1}\right) \\ &\quad \le H\left(K^{N-1} \mid Z^{N-1}\right) + I\left(M,U^{N-1};Y^{N-1} \mid K^{N-1}, Z^{N-1}\right) \\ &\qquad + I\left(U_N;Y_N \mid Z_N\right) + H\left(K_N\right) \\ &\quad \le H\left(K^{N-2} \mid Z^{N-2}\right) + I\left(M,U^{N-2};Y^{N-2} \mid K^{N-2}, Z^{N-2}\right) \\ &\qquad + I\left(U_{N-1};Y_{N-1} \mid Z_{N-1}\right) + H\left(K_{N-1}\right) \\ &\qquad + I\left(U_N;Y_N \mid Z_N\right) + H\left(K_N\right) \\ &\quad \le \sum_{i=1}^{N}\left(I\left(U_i;Y_i \mid Z_i\right) + H\left(K_i\right)\right) \\ &\quad \le \sum_{i=1}^{N}\left(I\left(U_i;Y_i \mid Z_i\right) + R_f\right). \end{aligned} \tag{B.14}$$

Substituting (B.14) into (B.13), we get

$$R_e \le \lim_{N\to\infty} \frac{1}{N}\left(\sum_{i=1}^{N}\left(I\left(U_i;Y_i \mid Z_i\right)+R_f\right)+\eta\left(P_e\right)\right). \tag{B.15}$$

To prove (B.8), consider

$$\begin{aligned} R_e &\le \lim_{N\to\infty}\Delta \\ &= \lim_{N\to\infty}\frac{1}{N}H\left(M \mid Z^N\right) \\ &=^{(b6)} \lim_{N\to\infty}\frac{1}{N}H\left(A^N \mid Z^N\right) \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}H\left(A_i \mid A^{i-1},Z^N\right) \\ &\le \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}H\left(A_i \mid Z_i\right), \end{aligned} \tag{B.16}$$

where $(b6)$ follows from the fact that the action sequence $a^N$ is a deterministic function of the message $m$.

To serve the single-letter characterization, let us introduce a time-sharing random variable $Q$ independent of $M$, $A^N, S^N, U^N, X^N, Y^N$, and $Z^N$ and uniformly distributed over $\{1,2,\ldots,N\}$. Set

$$\begin{gathered} U=\left(M,Y^{Q-1},S^{Q-1},Q\right), \\ A=A_Q, \qquad S=S_Q, \qquad X=X_Q, \\ Y=Y_Q, \qquad Z=Z_Q. \end{gathered} \tag{B.17}$$

It is straightforward to see that $(A,U)\to(X,S)\to Y\to Z$ and $U\to Y\to Z$ form Markov chains. Then, it is easy to get

$$\begin{aligned} I(U;Y\mid Z) &= H(U\mid Z)-H(U\mid YZ) \\ &= H(U)-H(U\mid YZ)-(H(U)-H(U\mid Z)) \\ &=^{(b7)} H(U)-H(U\mid Y)-(H(U)-H(U\mid Z)) \\ &= I(U;Y)-I(U;Z), \end{aligned} \tag{B.18}$$

where $(b7)$ follows from the Markov chain $U\to Y\to Z$. After using the standard time sharing argument [19, Section 5.4], utilizing (B.18), and letting $P_e\to 0$, (B.11), (B.15), and (B.16) are simplified into (B.6), (B.7), and (B.8), respectively.

This completes the proof of the converse part of Theorem 4.

## C. Coding Schemes for Noncausal Channel State Information

This section provides two different coding schemes for $H(A\mid Z)\ge\min\{I(U;Y)-I(U;S\mid A), I(U;Y)-I(U;Z)+R_f\}$ and $H(A\mid Z)\le\min\{I(U;Y)-I(U;S\mid A), I(U;Y)-I(U;Z)+R_f\}$ in Cases 1 and 2, respectively.

*Case 1* ($H(A\mid Z)\ge\min\{I(U;Y)-I(U;S\mid A), I(U;Y)-I(U;Z)+R_f\}$). In this case, we need to prove that all rate pairs $(R,R_e)$, satisfying

$$\begin{gathered} 0\le R_e\le R, \\ R\le I(U;Y)-I(U;S\mid A), \\ R_e\le I(U;Y)-I(U;Z)+R_f, \end{gathered} \tag{C.1}$$

are achievable. It is sufficient to prove the rate pairs $(R, R_e=\min\{I(U;Y)-I(U;S\mid A), I(U;Y)-I(U;Z)+R_f\})$ are achievable. In the following coding scheme, Gel'fand and Pinsker's coding technique [20] and rate splitting are used. The feedback symbol is used as a shared secret key between the transmitter and the receiver. The wiretapper has no knowledge of the secret key.

Define

$$\begin{gathered} R'=I(U;Y)-I(U;Z); \\ R'_f=R-R'. \end{gathered} \tag{C.2}$$

Split the message $\mathscr{M}$ into two parts, that is, $\mathscr{M}=(\mathscr{M}_1,\mathscr{M}_2)$. The corresponding random variables $M_1$, $M_2$ are uniformly distributed over $\mathscr{M}_1=\{1,2,3,\ldots,2^{NR'}\}$ and $\mathscr{M}_2=\{1,2,3,\ldots,2^{NR'_f}\}$. Since

$$\begin{aligned} R &\le C_{\mathrm{sn}}\left(R_f\right) \\ &= \min\left\{I(U;Y)-I(U;S\mid A), I(U;Y) \right. \\ &\qquad \left. -I(U;Z)+R_f, H(A\mid Z)\right\} \\ &= \min\left\{I(U;Y)-I(U;S\mid A), I(U;Y)-I(U;Z)+R_f\right\}, \end{aligned} \tag{C.3}$$

it is easy to get

$$\begin{aligned} R'_f &= R-R' \\ &\le \min\left\{I(U;Y)-I(U;S\mid A), I(U;Y)-I(U;Z)+R_f\right\} \\ &\quad -(I(U;Y)-I(U;Z)) \\ &= \min\left\{I(U;Z)-I(U;S\mid A), R_f\right\}. \end{aligned} \tag{C.4}$$

*Codebook Generation and Encoding.* Let $R=I(U;Y)-I(U;S\mid A)-\tau_1$, where $\tau_1$ is a fixed positive number. Generate $2^{NR'}$ i.i.d action sequences $a^N(m_1)$ according to $p(a^N)=\prod_{i=1}^{N}p(a_i)$ where $m_1\in\mathscr{M}_1$. The channel state sequence $s^N$ is generated through the discrete memoryless channel $p(s^N\mid a^N)=\prod_{i=1}^{N}p(s_i\mid a_i)$.

The feedback link is shared between the transmitter and the legitimate receiver. It is assumed that a secret key $k\in\{1,2,3,\ldots,2^{NR'_f}\}$ is transmitted through the feedback link and the wiretapper has no knowledge about the secret key. Let the corresponding random variable $K$ be independent of $M$ and uniformly distributed over $\{1,2,3,\ldots,2^{NR'_f}\}$.

For each $a^N(m_1)$, generate $2^{N(I(U;Y)-R'-\epsilon)}$ i.i.d sequences $u^N$ according to $p(u^N \mid a^N) = \prod_{i=1}^N p(u_i \mid a_i)$, where $\epsilon \to 0$ as $N \to \infty$. Note that

$$\begin{aligned} 2^{N(I(U;Y)-R'-\epsilon)} &= 2^{N(I(U;Y)-R+R'_f-\epsilon)} \\ &= 2^{N(I(U;S\mid A)+\tau_1+R'_f-\epsilon)}. \end{aligned} \tag{C.5}$$

Then, put these sequences into $2^{NR'_f}$ bins so that each bin contains $2^{N(I(U;S\mid A)+\tau_1-\epsilon)}$ sequences $u^N(m_1, t_b, t_u)$, where the bin index $t_b \in \{1, 2, 3, \ldots, 2^{NR'_f}\}$ and the sequence index $t_u \in \{1, 2, 3, \ldots, 2^{N(I(U;S\mid A)+\tau_1-\epsilon)}\}$.

Suppose that $k \in \{1, 2, 3, \ldots, 2^{NR'_f}\}$ is the shared secret key between the transmitter and the receiver during one transmission. To send message $m = (m_1, m_2)$ with the action sequence $a^N(m_1)$ and the state sequence $s^N$, where $m_1 \in \mathscr{M}_1$, $m_2 \in \mathscr{M}_2$, the transmitter tries to find a sequence $u^N(m_1, t_b, t_u)$ from the bin indexed by $t_b = m_2 \oplus k$ such that $(u^N(m_1, t_b, t_u), s^N, a^N(m_1)) \in T^N_{US|A}$. Here, $\oplus$ is modulo addition over $\{1, 2, 3, \ldots, 2^{NR'_f}\}$. If no such sequence exists in that bin, take $t_u = 1$. (In fact, at least one such sequence will exist because there are more than $2^{NI(U;S\mid A)}u^N(m_1, t_b, t_u)$ in each bin.) Then, The input sequence of the main channel is generated by $p(x^N \mid u^N, s^N) = \prod_{i=1}^N p(x_i \mid u_i, s_i)$.

*Decoding.* When observing the channel output $y^N$, the receiver tries to find a sequence $u^N(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u)$ such that $(u^N(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u), y^N) \in T^N_{UY}$. Then, the decoder outputs $(\widehat{m}_1, \widehat{t}_b \ominus k)$, where $\ominus$ is modulo subtraction over $\{1, 2, 3, \ldots, 2^{NR'_f}\}$. If no such $(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u)$ exists, take $(\widehat{m}_1, \widehat{t}_b, \widehat{t}_u) = (1, 1, 1)$.

*Analysis of Error Probability.* By using similar arguments in [1], it is easy to show that the legitimate receiver can decode the message with vanishing decoding error probability. Note that since Gel'fand and Pinsker's coding technique is used for the noncausal case, particular attention should be paid to the encoding error probability when calculating the whole error probability. The details for proving $P_e \le \epsilon$ are omitted.

*Analysis of Equivocation.* We focus on calculating wiretapper's equivocation of the message, that is, the uncertainty about the message given wiretapper's observation.

Using the same derivation in (A.6), we have

$$H\left(M \mid Z^N\right) \ge H\left(M_1 \mid Z^N\right) + NR'_f. \tag{C.6}$$

The first term in (C.6) is calculated by applying Csiszar's method for analyzing equivocation [8]:

$$\begin{aligned} &H\left(M_1 \mid Z^N\right) \\ &\quad = H\left(M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad = H\left(M_1, Z^N, U^N\right) - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad = H\left(M_1, U^N\right) + H\left(Z^N \mid M_1, U^N\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad \ge H\left(U^N\right) + H\left(Z^N \mid M_1, U^N\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad \ge H\left(U^N\right) - H\left(U^N \mid Y^N\right) + H\left(Z^N \mid M_1, U^N\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad = I\left(U^N; Y^N\right) + H\left(Z^N \mid M_1, U^N\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad =^{(c1)} I\left(U^N; Y^N\right) + H\left(Z^N \mid U^N\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad =^{(c2)} NI\left(U; Y\right) + NH\left(Z \mid U\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - H\left(Z^N\right) \\ &\quad \ge^{(c3)} NI\left(U; Y\right) + NH\left(Z \mid U\right) \\ &\qquad - H\left(U^N \mid M_1, Z^N\right) - NH\left(Z\right) \\ &\quad \ge^{(c4)} NI\left(U; Y\right) - NI\left(U; Z\right) - N\epsilon_3. \end{aligned} \tag{C.7}$$

In the above derivations, $(c1)$ follows from the fact that $M_1 \to U^N \to Z^N$ is a Markov chain. $(c2)$ follows from the fact that $S^N$, $U^N$, and $X^N$ are i.i.d generated and both the main channel and wiretap channel are discrete memoryless. $(c3)$ follows from $H(Z^N) = \sum_{i=1}^N H(Z_i \mid Z^{i-1}) \le NH(Z)$, where $Z$ is regarded as the general random variable of $Z_i$. To verify $(c4)$, note the fact that given the message $m_1$, the number of $U^N$ is

$$\begin{aligned} 2^{N(I(U;Y)-R'-\epsilon)} &= 2^{N[I(U;Y)-(I(U;Y)-I(U;Z))-\epsilon]} \\ &= 2^{N(I(U;Z)-\epsilon)} \le 2^{NI(U;Z)}. \end{aligned} \tag{C.8}$$

By applying the standard channel coding theorem [23, theorem 7.7.1], the wiretapper can decode $U^N$ with vanishing decoding error probability. Then, utilizing Fano's inequality, we get $H(U^N \mid M_1, Z^N) \le N\epsilon_3$, where $\epsilon_2 \to 0$ when $N \to \infty$. (C.7) is verified.

Substituting (C.7) into (C.6) and using the same derivation in (A.9), the achievability proof of Case 1 is completed.

*Case 2* ($H(A \mid Z) \le \min\{I(U;Y) - I(U;S \mid A), I(U;Y) - I(U;Z) + R_f\}$). In this case, we show that all rate pairs $(R, R_e)$, satisfying

$$0 \le R_e \le R,$$
$$R \le I(U;Y) - I(U;S \mid A), \tag{C.9}$$
$$R_e \le H(A \mid Z),$$

are achievable. It is sufficient to prove that the rate pairs $(R, R_e = \min\{I(U;Y) - I(U;S \mid A), H(A \mid Z)\})$ are achievable. Gel'fand and Pinsker's coding technique is used. The proof is similar to that in [18].

*Codebook Generation and Encoding.* Let $R = I(U;Y) - I(U;S \mid A) - \tau_2$, where $\tau_2$ is a fixed positive number. $M$ is uniformly distributed over $\mathscr{M} = \{1, 2, 3, \ldots, 2^{NR}\}$. Generate $2^{NR}$ action sequences $a^N(m)$ according to $p(a^N) = \prod_{i=1}^N p(a_i)$, where $m \in \mathscr{M}$. The channel state sequence $s^N$ is generated through the discrete memoryless channel $p(s^N \mid a^N) = \prod_{i=1}^N p(s_i \mid a_i)$. For each $a^N(m)$, generate $2^{N(I(U;Y)-R-\epsilon)} = 2^{N(I(U;S \mid A)+\tau_2-\epsilon)}$ i.i.d sequences $u^N(m,t)$ according to $p(u^N \mid a^N) = \prod_{i=1}^N p(u_i \mid a_i)$, where $t \in \{1, 2, 3, \ldots, 2^{N(I(U;S \mid A)+\tau_2-\epsilon)}\}$.

To send message $m$ with the action sequence $a^N(m)$ and the state sequence $s^N$, the transmitter chooses a sequence $u^N(m,t)$ such that $(u^N(m,t), s^N, a^N(m)) \in T^N_{US \mid A}$. If no such sequence exists, take $t = 1$. Then, the input sequence of the main channel is generated by $p(x^N \mid u^N, s^N) = \prod_{i=1}^N p(x_i \mid u_i, s_i)$.

*Decoding.* When observing the channel output $y^N$, the receiver tries to find a sequence $u^N(\widehat{m}, \widehat{t})$ such that $(u^N(\widehat{m}, \widehat{t}), y^N) \in T^N_{UY}$. Then, the decoder outputs $\widehat{m}$. If no such $\widehat{m}$ exists, take $\widehat{m} = 1$.

*Analysis of Error Probability.* By using similar arguments in [1], it is easy to show that the legitimate receiver can decode the message with vanishing decoding error probability. Note that since Gel'fand and Pinsker's coding technique is used for the noncausal case, particular attention should be paid to the encoding error probability when calculating the whole error probability. The details for proving $P_e \le \epsilon$ are omitted.

The *analysis of equivocation* is the same as (A.11). The achievability proof of Case 2 is finished.

The achievability proof of Theorem 5 is completed.

## D. Proof of the Converse Part of Theorem 5

This section proves that all achievable rate pairs $(R, R_e)$ for the channel model in Figure 1 with noncausal states are contained in $\mathscr{R}_n$. Concretely, we need to show that, for any rate pair $(R, R_e)$, feedback alphabet $\mathscr{K}$, and decoding error probability $P_e$ satisfying (2), there exist random variables $(A, U) \to (X, S) \to Y \to Z$ such that

$$0 \le R_e \le R, \tag{D.1}$$
$$R \le I(U;Y) - I(U;S \mid A), \tag{D.2}$$
$$R_e \le I(U;Y) - I(U;Z) + R_f, \tag{D.3}$$
$$R_e \le H(A \mid Z). \tag{D.4}$$

The four inequalities (D.1), (D.2), (D.3), and (D.4) are proved as follows.

(D.1), (D.3), and (D.4) can be proved using the same derivations in (B.9), (B.15), and (B.16), respectively, except for the identification of the auxiliary random variable $U$. Therefore, we focus on proving (D.2) and giving the identification of the auxiliary random variable $U$.

A lemma in [13] is presented first.

**Lemma 6** (see [13]). *One has*

$$\sum_{i=1}^N I\left(S_{i+1}^N, A^N; Y_i \mid M, Y^{i-1}\right) = \sum_{i=1}^N I\left(S_i; Y^{i-1} \mid M, S_{i+1}^N, A^N\right). \tag{D.5}$$

*Proof.* One has

$$\begin{aligned}
&\sum_{i=1}^N I\left(S_{i+1}^N, A^N; Y_i \mid M, Y^{i-1}\right) \\
&\quad =^{(d1)} \sum_{i=1}^N I\left(S_{i+1}^N, A^N; Y_i \mid M, Y^{i-1}, A^N\right) \\
&\quad = \sum_{i=1}^N I\left(S_{i+1}^N; Y_i \mid M, Y^{i-1}, A^N\right) \\
&\quad = \sum_{i=1}^N \sum_{j=i+1}^N I\left(S_j; Y_i \mid M, S_{j+1}^N, Y^{i-1}, A^N\right) \\
&\quad = \sum_{j=1}^N \sum_{i=j+1}^N I\left(S_i; Y_j \mid M, S_{i+1}^N, Y^{j-1}, A^N\right) \\
&\quad = \sum_{i=1}^N \sum_{j=1}^{i-1} I\left(S_i; Y_j \mid M, S_{i+1}^N, Y^{j-1}, A^N\right) \\
&\quad = \sum_{i=1}^N I\left(S_i; Y^{i-1} \mid M, S_{i+1}^N, A^N\right),
\end{aligned} \tag{D.6}$$

where $(d1)$ follows from the fact that the action sequence is a deterministic function of the transmitted message.

Since the transmission rate is

$$R = \lim_{N \to \infty} \frac{1}{N} \log \|\mathscr{M}\| = \lim_{N \to \infty} \frac{1}{N} H(M), \tag{D.7}$$

let us consider

$$\begin{aligned}
&\frac{1}{N} H(M) \\
&\quad = \frac{1}{N} I\left(M; Y^N\right) + \frac{1}{N} H\left(M \mid Y^N\right) \\
&\quad \le^{(d2)} \frac{1}{N} I\left(M; Y^N\right) + \frac{1}{N} \eta\left(P_e\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{i=1}^{N} I\left(M; Y_i \mid Y^{i-1}\right) + \frac{1}{N}\eta\left(P_e\right) \\
&\leq \frac{1}{N}\sum_{i=1}^{N} I\left(M, Y^{i-1}; Y_i\right) + \frac{1}{N}\eta\left(P_e\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[I\left(M, Y^{i-1}, S_{i+1}^N, A^N; Y_i\right)\right. \\
&\qquad \left. - I\left(S_{i+1}^N, A^N; Y_i \mid M, Y^{i-1}\right)\right] + \frac{1}{N}\eta\left(P_e\right) \\
&\overset{(d3)}{=} \frac{1}{N}\sum_{i=1}^{N} \left[I\left(M, Y^{i-1}, S_{i+1}^N, A^N; Y_i\right)\right. \\
&\qquad \left. - I\left(S_i; Y^{i-1} \mid M, S_{i+1}^N, A^N\right)\right] + \frac{1}{N}\eta\left(P_e\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[I\left(M, Y^{i-1}, S_{i+1}^N, A^N; Y_i\right)\right. \\
&\qquad \left. - I\left(S_i; Y^{i-1} \mid M, S_{i+1}^N, A^N, A_i\right)\right] + \frac{1}{N}\eta\left(P_e\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[I\left(M, Y^{i-1}, S_{i+1}^N, A^N; Y_i\right)\right. \\
&\qquad - I\left(S_i; M, Y^{i-1}, S_{i+1}^N, A^N \mid A_i\right) \\
&\qquad \left. + I\left(S_i; M, S_{i+1}^N, A^N \mid A_i\right)\right] + \frac{1}{N}\eta\left(P_e\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left[I\left(M, Y^{i-1}, S_{i+1}^N, A^N; Y_i\right)\right. \\
&\qquad - I\left(S_i; M, Y^{i-1}, S_{i+1}^N, A^N \mid A_i\right) \\
&\qquad \left. + I\left(S_i; M, S_{i+1}^N, A^{i-1}, A_{i+1}^N \mid A_i\right)\right] + \frac{1}{N}\eta\left(P_e\right) \\
&\overset{(d4)}{=} \frac{1}{N}\sum_{i=1}^{N} \left[I\left(M, Y^{i-1}, S_{i+1}^N, A^N; Y_i\right)\right. \\
&\qquad \left. - I\left(S_i; M, Y^{i-1}, S_{i+1}^N, A^N \mid A_i\right)\right] + \frac{1}{N}\eta\left(P_e\right) \\
&\overset{(d5)}{=} \frac{1}{N}\sum_{i=1}^{N} \left[I\left(U_i; Y_i\right) - I\left(S_i; U_i \mid A_i\right)\right] + \frac{1}{N}\eta\left(P_e\right),
\end{aligned}
\tag{D.8}
$$

where $(d2)$ follows from Fano's inequality, $(d3)$ is from Lemma 6, $(d4)$ follows from the fact that $S_i \to A_i \to (M, S_{i+1}^N, A^{i-1}, A_{i+1}^N)$ form a Markov chain, and $(d5)$ follows from defining $U_i = (M, Y^{i-1}, S_{i+1}^N, A^N)$. Substituting (D.8) into (D.7),

$$
\begin{aligned}
R &= \lim_{N\to\infty} \frac{1}{N} H(M) \\
&\leq \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \left[I\left(U_i; Y_i\right) - I\left(S_i; U_i \mid A_i\right)\right].
\end{aligned}
\tag{D.9}
$$

To serve the single-letter characterization, let us introduce a time-sharing random variable $Q$ independent of $M$, $A^N$, $S^N$, $U^N$, $X^N$, $Y^N$, and $Z^N$ and uniformly distributed over $\{1, 2, \ldots, N\}$. Set

$$
\begin{gathered}
U = \left(M, Y^{Q-1}, S_{Q+1}^N, A^N, Q\right), \\
A = A_Q, \qquad S = S_Q, \qquad X = X_Q, \\
Y = Y_Q, \qquad Z = Z_Q.
\end{gathered}
\tag{D.10}
$$

It is straightforward to see that $(A, U) \to (X, S) \to Y \to Z$ form a Markov chain. Applying the standard time-sharing argument [19, Section 5.4] to (D.9), we can easily get (D.2).

This completes the proof of the converse part of Theorem 5. □

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants no. 61171173, 60932003, and 61271220. The authors also would like to thank the anonymous reviewers for helpful comments.

## References

[1] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5396–5411, 2010.

[2] B. Ahmadi and O. Simeone, "On channels with action-dependent states," in *Proceedings of the IEEE Information Theory Workshop (ITW '12)*, Lausanne, Switzerland, September 2012.

[3] Y. Steinberg and T. Weissman, "The degraded broadcast channel with action-dependent states," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '12)*, Boston, Mass, USA, July 2012.

[4] H. Asnani, H. Permuter, and T. Weissman, "Probing capacity," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7317–7332, 2011.

[5] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Coding With action-dependent side information and additional reconstruction requirements," Submitted to *IEEE Transactions on Information Theory*, http://arxiv.org/abs/1202.1484v1.

[6] C. Choudhuri and U. Mitra, "Action dependent strictly causal state communication," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '12)*, Cambridge, Mass, USA, July 2012.

[7] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[8] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[9] A. Khisti, G. Wornell, A. Wiesel, and Y. Eldar, "On the Gaussian MIMO wiretap channel," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 2471–2475, Nice, France, June 2007.

[10] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: the 2-2-1 channel," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4033–4039, 2009.

[11] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[12] B. Dai and Y. Luo, "Some new results on the wiretap channel with side information," *Entropy*, vol. 14, no. 9, pp. 1671–1702, 2012.

[13] B. Dai, A. J. H. Vinck, Y. Luo, and X. Tang, "Wiretap channel with action-dependent channel state information," *Entropy*, vol. 15, no. 2, pp. 445–473, 2013.

[14] E. Ekrem and S. Ulukus, "Multi-receiver wiretap channel with public and confidential messages," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2165–2177, 2013.

[15] R. Ahlswede and N. Cai, "Transmission, identification and common randomness capacities for wire-tape channels with secure feedback from the decoder," in *General Theory of Information Transfer and Combinatorics*, vol. 4123 of *Lecture Notes in Computer Science*, pp. 258–275, Springer, Berlin, Germany, 2006.

[16] E. Ardestanizadeh, M. Franceschetti, T. Javidi, and Y.-H. Kim, "Wiretap channel with secure rate-limited feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5353–5361, 2009.

[17] X. Yin, Z. Xue, and B. Dai, "Capacity-equivocation regions of the DMBCs with noiseless feedback," *Mathematical Problems in Engineering*, vol. 2013, Article ID 102069, 13 pages, 2013.

[18] B. Dai, A. J. H. Vinck, and Y. Luo, "Wiretap channel in the presence of action-dependent states and noiseless feedback," *Journal of Applied Mathematics*, vol. 2013, Article ID 423619, 17 pages, 2013.

[19] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, Cambridge, UK, 2011.

[20] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[21] I. Csiszár and J. Körner, *Information Theory*, Probability and Mathematical Statistics, Academic Press, london, UK, 1981.

[22] C. Choudhuri and U. Mitra, "How useful is adaptive action?" in *Proceedings of the Global Communications Conference (GLOBECOM '12)*, Anaheim, Calif, USA, December 2012.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.