Entropy **2014**, *16*, 2105-2130; doi:10.3390/e16042105

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Wiretap Channel with Information Embedding on Actions

Xinxing Yin * and Zhi Xue

Electronic Engineering Department, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai 200240, China; E-Mail: zxue@sjtu.edu.cn

* Author to whom correspondence should be addressed; E-Mail: yinxinxing@sjtu.edu.cn; Tel.: +86-21-34205982.

Received: 11 November 2013; in revised form: 28 March 2014 / Accepted: 10 April 2014 / Published: 14 April 2014

Abstract: Information embedding on actions is a new channel model in which a specific decoder is used to observe the actions taken by the encoder and retrieve part of the message intended for the receiver. We revisit this model and consider a different scenario where a secrecy constraint is imposed. By adding a wiretapper in the model, we aim to send the confidential message to the receiver and keep it secret from the wiretapper as much as possible. We characterize the inner and outer bounds on the capacity-equivocation region of such a channel with noncausal (and causal) channel state information. Furthermore, the lower and upper bounds on the sum secrecy capacity are also obtained. Besides, by eliminating the specific decoder, we get a new outer bound on the capacity-equivocation region of the wiretap channel with action-dependent states and prove it is tighter than the existing outer bound. A binary example is presented to illustrate the tradeoff between the sum secrecy rate and the information embedding rate under the secrecy constraint. We find that the secrecy constraint and the communication requirements of information embedding have a negative impact on improving the secrecy transmission rate of the given communication link.

Keywords: information embedding; wiretap channel; action-dependent states; sum secrecy capacity

---

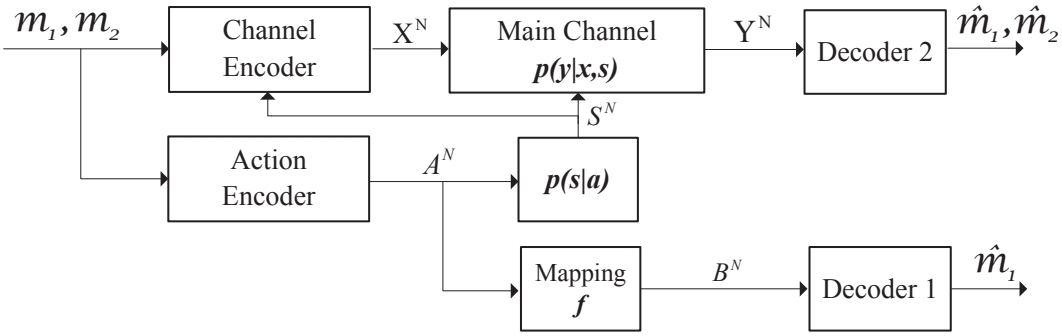
1. Introduction

In wireless communication systems, such as sensor networks, mobile networks and satellite communications, sensitive data is transferred through multiple hops. The nodes in the network usually need to take various type of actions to acquire the state of the network before transferring the packets. The state acquisition in the network often requires the exchange of control information, which uses physical resources within the system. For example, routers measure the network congestion levels via the transmission of probing packets; wireless transceivers evaluate the channel quality through training or feedback; and radio terminals switch among different operating modes, such as transmit, receive or idle. Based on these motivating observations, Weissman introduced channels with action-dependent states where the encoder in a point-to-point channel could take actions to affect the channel state information [1]. The capacity of such a channel where the channel inputs depended noncausally (and causally) on the channel states was determined. After Weissman's publication, [2] investigated the channel model where both the channel encoder and decoder could probe the channel states and obtained the cost constrained "probing capacity". Different from this, [1–3] studied the degraded broadcast channel with causal action-dependent states where the transmitter sent two kinds of messages to two different receivers. The capacity region was derived. For the noncausal case, only inner and outer bounds on the capacity region were obtained [4]. Other extensions of the channel with action-dependent states can be seen in [5–10].

Recently, [11,12] explored information embedding on actions in the channel with noncausal action-dependent states; see Figure 1. In this new setup, an additional decoder was introduced to observe a function of the actions taken by the encoder. It tried to get part of the transmitted message. Actually, the idea of "information embedding" on actions in such a channel is related to the classical topic of information hiding (e.g., [13–17]) and could be explained by the following example. In communication networks, probing the congestion state requires sending training packets to the nearest router. Meanwhile, the router (the "recipient" of the actions) may need to obtain partial information, such as the header of the packet, to find the address of the intended receiver. Since the actions play the role of providing necessary information about the message for the router, it is natural to ask how much information could be embedded in the actions without affecting the system performance. [11] got the capacity-cost region and showed that the communication requirements of the action-cribbing decoder were generally in conflict with the goal of improving the efficiency of the communication link.

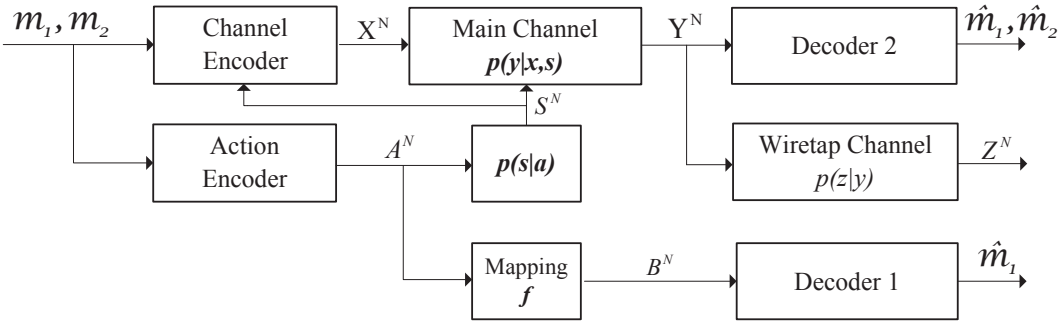
However, the above action-dependent channel models [1–12] considered no secrecy constraint, which was extremely important in communications. For instance, the broadcast nature of wireless networks gives rise to the hidden danger of information leakage to malicious receiver when broadcasting the sensitive data and acquiring the state information. Recent works [18,19] studied the secure communication problems in channels with action-dependent states. [18] added a wiretapper to the model in [1] and got the inner and outer 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. [19] studied the effects of feedback on the secrecy capacity, which is the maximum rate of data transmission at which the message can be communicated in perfect secrecy, of the wiretap channel with action-dependent states.

Figure 1. Information embedding on actions [11].



From the perspective of secure communication, we consider a different communication model in Figure 2, *i.e.*, the wiretap channel with information embedding on actions. In this setup, the transmitter aims to send the confidential message to the receiver and keep it secret from the wiretapper as much as possible. We use equivocation (*i.e.*, the uncertainty about the confidential message) at the wiretapper to measure the level of information leakage. Meanwhile, like [11], a specific decoder is introduced to retrieve a portion of the confidential message (see m_1 in Figure 2). The specific decoder observes a function of the message-dependent actions, which affects the formulation of the channel states. Our work is novel in the sense that we consider the secrecy constraint in the information transmission, and we try to characterize how much information could be embedded in the actions without increasing the information leakage.

Figure 2. Wiretap channel with information embedding on actions.



For the new channel model described above, this paper obtains the inner and outer bounds on the capacity-equivocation region of such a channel with noncausal (and causal) channel state information. Furthermore, the lower and upper bounds on the sum secrecy capacity are also attained. Through a special case where no message needs to be retrieved by the specific decoder, we get a new outer bound on the capacity-equivocation region of the wiretap channel with action-dependent states and prove that it is tighter than the existing outer bound. To illustrate the tradeoff between the sum secrecy rate and the information embedding rate under the secrecy constraint, we provide a binary example. It shows that the sum secrecy rate is reduced when the information embedding rate increases. We find that the secrecy constraint and the communication requirements of the specific decoder have a negative impact on improving the secrecy transmission rate of the given communication link.

The rest of the article is organized as follows. Section 2 describes the wiretap channel with information embedding on actions and outlines the main results. Section 3 discusses the results and presents a binary example. We conclude in Section 4 with a summary of the whole work and some future directions.

2. Channel Model and Main Results

The symbol notations and description of the channel model are presented in Subsection 2.1. Subsection 2.2 characterizes the inner and outer bounds on the capacity-equivocation region of the wiretap channel with information embedding on actions.

2.1. Symbol Notations and Channel Model

Throughout this paper, we use calligraphic letters, e.g., $\mathcal{X}$, $\mathcal{Y}$, to denote the finite sets and $\|\mathcal{X}\|$ to denote the cardinality of the set, $\mathcal{X}$. Uppercase letters, e.g., X, Y, are used to denote random variables taking values from finite sets, e.g., $\mathcal{X}$, $\mathcal{Y}$. 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 $(Z_i, Z_{i+1}, ..., Z_j)$ of random variables for $1 \leq i \leq j$ and 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 $0 \leq \epsilon \leq 1$, the set of the typical N-sequences x^N is defined as $T_X^N(\epsilon) = \{x^N : |\pi(x|x^N) - p(x)| \leq \epsilon p(x) \text{ for all } x \in \mathcal{X}\}$, where $\pi(x|x^N)$ denotes the frequency of occurrences of letter x in the sequence, x^N (for more details about typical sequences, please refer to [23,24]). The set of the conditional typical sequences, e.g., $T_{Y|X}^N(\epsilon)$, follows similarly. In this paper, it is assumed that the base of the log function is two.

The wiretap channel with information embedding on actions is depicted in Figure 2. We aim to send the confidential message (M_1, M_2) to the legitimate receiver through such a channel and keep it secret from the wiretapper as much as possible. Part of the message embedded in the actions needs to be retrieved by the specific decoder. We use equivocation at the wiretapper to measure the secrecy of the confidential message.

Concretely, the model of wiretap channel with information embedding on actions is specified by $\{\mathcal{A}, \mathcal{B}, \mathcal{S}, f, p(s|a), \mathcal{X}, p(y, z|s, x), \mathcal{Y}, \mathcal{Z}\}$. To send the message (M_1, M_2), an action sequence, $A^N(M_1, M_2)$, is first selected by the encoder. Then, the generation of the channel states, S^N, is affected by the actions, instead of by nature. The channel states, S^N, are generated through a discrete memoryless channel (DMC) $p(s^N|a^N(m_1, m_2)) = \prod_{i=1}^{N} p(s_i|a_i)$. The stochastic channel encoder, φ, is specified by a matrix of conditional probability distributions, $\varphi(x^N|m_1, m_2, s^N)$. Note that $\sum_{x^N} \varphi(x^N|m_1, m_2, s^N) = 1$, and $\varphi(x^N|m_1, m_2, s^N)$ is the probability that the message (m_1, m_2) and the state sequence, s^N, are encoded as the channel input, x^N. When the state sequence, s^N, is known causally by the channel encoder, the channel encoder at time i is specified by $\varphi_i(x_i|m_1, m_2, s^i)$, where x_i is the output of the channel encoder at time i and $s^i = (s_1, s_2, ..., s_i)$ is the channel states up to time i. When the channel encoder knows the state, s^N, in a noncausal manner, the channel encoder at time i is specified by $\varphi_i(x_i|m_1, m_2, s^N)$.

The main channel is a DMC with discrete input alphabet $\mathcal{X} \times \mathcal{S}$ and output alphabet $\mathcal{Y}$. The channel is memoryless in the sense that $p(y^N|x^N, s^N) = \prod_{i=1}^{N} p(y_i|x_i, s_i)$, where $y^N \in \mathcal{Y}^N$, $x^N \in \mathcal{X}^N$ and $s^N \in \mathcal{S}^N$. Decoder 1 observes signal B^N as a deterministic function of the actions, A^N, *i.e.*, $B^N = f(A^N)$. It estimates part of the transmitted message. Decoder 1 is specified by $\psi_1 : \mathcal{B}^N \to \mathcal{M}_1$. The output of Decoder 1 is $\hat{M}_1$. The probability of the error of Decoder 1 is defined as $P_{e1} = Pr\{\hat{M}_1 \neq M_1\}$. The legitimate receiver decodes the message $(\hat{\hat{M}}_1, \hat{\hat{M}}_2)$ by Decoder 2 (see Figure 2). Decoder 2 is specified by $\psi_2 : \mathcal{Y}^N \to \mathcal{M}_1 \times \mathcal{M}_2$. The probability of the error of Decoder 2 is defined as $P_{e2} = Pr\{(\hat{\hat{M}}_1, \hat{\hat{M}}_2) \neq (M_1, M_2)\}$. The wiretap channel is also a DMC with transition probability $p(z^N|y^N) = \prod_{i=1}^{N} p(z_i|y_i)$, where $z^N \in \mathcal{Z}^N$ is the observation of the wiretapper. The uncertainty of the message for the wiretapper is measured by $\lim_{N\to\infty} \Delta = \lim_{N\to\infty} \frac{H(M_1,M_2|Z^N)}{N}$. In our model, the wiretap channel is assumed to be degraded from the main channel, *i.e.*, $X \to Y \to Z$ form a Markov chain.

Then, we give the definition of "achievable" and "sum secrecy capacity" as follows.

Definition 1 *A rate triple (R_1, R_2, R_e) is said to be achievable for the model in Figure 2 if there exists a channel encoder-decoder, such that:*

$$\lim_{N\to\infty} \frac{log\|\mathcal{M}_1\|}{N} = R_1 \tag{1}$$

$$\lim_{N\to\infty} \frac{log\|\mathcal{M}_2\|}{N} = R_2 \tag{2}$$

$$\lim_{N\to\infty} \frac{H(M_1, M_2|Z^N)}{N} \geq R_e \tag{3}$$

$$P_{e1} \leq \epsilon, P_{e2} \leq \epsilon \tag{4}$$

where ϵ is an arbitrary small positive real number, (R_1, R_2) are the rates of the message (M_1, M_2) and R_e is the rate of equivocation. The capacity-equivocation region is defined as the convex closure of all achievable rate triples (R_1, R_2, R_e).

Definition 2 *The sum secrecy capacity is the maximum rate at which the confidential message can be sent to the receiver in perfect secrecy. The sum secrecy capacity:*

$$C_s = \max_{(R_1,R_2,R_e=R_1+R_2)\in\mathcal{R}} (R_1 + R_2) \tag{5}$$

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

Based on the definition in Equation (5), the sum secrecy capacity for the model in Figure 2 with noncausal action-dependent states is $C_{sn} = \max_{(R_1,R_2,R_e=R_1+R_2)\in\mathcal{R}_n} (R_1 + R_2)$, where $\mathcal{R}_n$ is the capacity-equivocation region of the noncausal case. Similarly, we can define the sum secrecy capacity of the causal case, C_{sc}.

2.2. Main Results

In this subsection, four theorems are presented. Theorems 1 and 2 give the inner and outer bounds on the capacity-equivocation region for the channel model in Figure 2 with noncausal action-dependent states. For the causal case, the inner and outer bounds are characterized in Theorems 3 and 4.

Theorem 1 *An achievable rate-equivocation region of the wiretap channel with information embedding on actions when the states are noncausally known to the channel encoder is the set:*

$$\begin{aligned}
&\mathcal{R}_{in} = \{(R_1, R_2, R_e) \\
&R_1 \leq H(B) && (6)\\
&R_1 + R_2 \leq I(U;Y) - I(U;S|A) && (7)\\
&R_e \leq R_1 + R_2 && (8)\\
&R_e \leq I(U;Y) - \max\{I(U;Z), I(U;S|A)\} && (9)\\
&R_e \leq H(A|Z)\} && (10)
\end{aligned}$$

where the joint distributions $p(a,b,u,x,s,y,z) = p(z|y)p(y|x,s)p(x|u,s)p(s|u,a)p(u|a)p(a)1_{\{b=f(a)\}}$, *which indicates* $(A,B,U) \to (X,S) \to Y \to Z$ *form a Markov chain.*

Theorem 2 *An outer bound on the capacity-equivocation region of the wiretap channel with information embedding on actions when the states are noncausally known to the channel encoder is the set:*

$$\begin{aligned}
&\mathcal{R}_{on} = \{(R_1, R_2, R_e) \\
&R_1 \leq H(B) && (11)\\
&R_1 + R_2 \leq I(U;Y) - I(U;S|A) && (12)\\
&R_e \leq R_1 + R_2 && (13)\\
&R_e \leq I(U;Y) - I(V;Z|Q) - I(U;S|V)\} && (14)
\end{aligned}$$

where the joint distributions $p(a,b,u,x,s,y,z) = p(z|y)p(y|x,s)p(x|u,s)p(q|v)p(v|u)p(a,u,s)1_{\{b=f(a)\}}$, *which indicates* $(B,A,U,V,Q) \to (X,S) \to Y \to Z$ *and* $Q \to V \to U \to Y \to Z$ *form Markov chains.*

Comments:

- Theorems 1 and 2 are proven in Appendix A.
- To exhaust $\mathcal{R}_{in}$ and $\mathcal{R}_{on}$, it is enough to restrict $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{Q}$ to satisfy:

$$\begin{aligned}
&\|\mathcal{U}\| \leq \|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\| + 2\\
&\|\mathcal{Q}\| \leq \|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\|\\
&\|\mathcal{V}\| \leq \|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\|(\|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\| + 1)
\end{aligned}$$

This can be easily proven by using the support lemma (see p. 310 in [25]).

Theorem 3 *An achievable rate-equivocation of the wiretap channel with information embedding on actions when the states are causally known to the channel encoder is the set:*

$$\begin{aligned}
&\mathcal{R}_{ic} = \{(R_1, R_2, R_e) \\
&R_1 \leq H(B)\\
&R_1 + R_2 \leq I(U;Y)\\
&R_e \leq R_1 + R_2\\
&R_e \leq I(U;Y) - I(U;Z)\\
&R_e \leq H(A|Z)\} && (15)
\end{aligned}$$

where the joint distributions $p(a,b,u,x,s,y,z) = p(z|y)p(y|x,s)p(x|u,s)p(s|a)p(u|a)p(a)1_{\{b=f(a)\}}$, *which indicates* $(A,B,U) \to (X,S) \to Y \to Z$ *and* $U \to A \to S$ *form Markov chains.*

Theorem 4 *An outer bound on the capacity-equivocation region of the wiretap channel with information embedding on actions when the states are causally known to the channel encoder is the set:*

$$\begin{aligned}
&\mathcal{R}_{oc} = \{(R_1, R_2, R_e) \\
&R_1 \leq H(B) && (16)\\
&R_1 + R_2 \leq I(U;Y) && (17)\\
&R_e \leq R_1 + R_2 && (18)\\
&R_e \leq I(U;Y) - I(V;Z|Q)\} && (19)
\end{aligned}$$

where the joint distributions $p(a,b,u,x,s,y,z) = p(z|y)p(y|x,s)p(x|u,s)p(q|v)p(v|u)p(a,u,s)1_{\{b=f(a)\}}$, *which indicates* $(B,A,U,V,Q) \to (X,S) \to Y \to Z$ *and* $Q \to V \to U \to Y \to Z$ *form Markov chains.*

Comments:

- Theorems 3 and 4 are proven in Appendix B.
- To exhaust $\mathcal{R}_{ic}$ and $\mathcal{R}_{oc}$, it is enough to restrict $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{Q}$ to satisfy:

$$\begin{aligned}
\|\mathcal{U}\| &\leq \|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\| + 1\\
\|\mathcal{Q}\| &\leq \|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\|\\
\|\mathcal{V}\| &\leq (\|\mathcal{A}\|\|\mathcal{X}\|\|\mathcal{S}\|)^2
\end{aligned}$$

This can be easily proven by using the support lemma (see p. 310 in [25]).

Further discussion about the theorems and the comparison with other existing results are given in Section 3.

3. Discussion and Example

In this section, we first calculate the sum secrecy capacities of the noncausal and causal cases. Then, we compare our results with some existing results and present a binary example to illustrate the tradeoff between the sum secrecy rate and the information embedding rate under the secrecy constraint.

3.1. Discussion

Corollary 1 *The lower and upper bounds on the sum secrecy capacity of the model in Figure 2 with noncausal action-dependent states are:*

$$C_{ln} = \max_{p(x|u,s)p(u|s,a)p(a)} \min\{I(U;Y) - \max\{I(U;Z), I(U;S|A)\}, H(A|Z)\} \quad (20)$$

and:

$$C_{un} = \max_{p(x|u,s)p(q|v)p(v|u)p(u|s,a)p(a)} \min\{I(U;Y) - I(U;S|A), I(U;Y) - I(V;Z|Q) - I(U;S|V)\} \quad (21)$$

respectively.

Proof: *According to the definition of Formula (5), $\mathcal{R}_{in}$ and $\mathcal{R}_{on}$, we can easily get Equations (20) and (21).*

Similarly, we have the following corollary for the causal case.

Corollary 2 *The lower and upper bounds on the sum secrecy capacity of the model in Figure 2 with causal action-dependent states are:*

$$C_{lc} = \max_{p(u,a)p(x|u,s)} \min\{I(U;Y) - I(U;Z), H(A|Z)\} \tag{22}$$

and:

$$C_{uc} = \max_{p(x|s,u)p(q|v)p(v|u)p(u|a)p(a)} I(U;Y) - I(V;Z|Q) \tag{23}$$

respectively.

Proof: *According to the definition of Formula (5), $\mathcal{R}_{ic}$ and $\mathcal{R}_{oc}$, we can easily get Equations (22) and (23).*

According to [11], the capacity region of the model in Figure 1 (without the secrecy constraint) is:

$$\begin{aligned} &\mathcal{R}_E = \{(R_1, R_2) \\ &R_1 \leq H(B) &(24)\\ &R_1 + R_2 \leq I(U;Y) - I(U;S|A)\} &(25)\end{aligned}$$

Then, the corresponding capacity is:

$$C_E = \max_{p(u,a,x,s)} I(U;Y) - I(U;S|A) \tag{26}$$

Compare Formula (26) with (20); we can get $C_{ln} \leq C_E$. This implies that the secrecy constraint reduces the secrecy transmission rate of the communication link. Therefore, once the problem of information leakage is considered, the system designer has to trade off between the transmission rate and data security. Moreover, without the secrecy constraint, we have the following corollary.

Corollary 3 *Without considering the secrecy constraint* (i.e., *ignoring R_e in $\mathcal{R}_{in}$), we arrive at the results in [11].*

Proof: *This corollary is verified by setting $R_e = 0$ in $\mathcal{R}_{in}$.*

When no message needs to be embedded in the actions, the model in Figure 2 turns to the wiretap channel with action-dependent states [18]; see Figure 3. [18] gave an upper bound on the secrecy capacity of the model with noncausal states as:

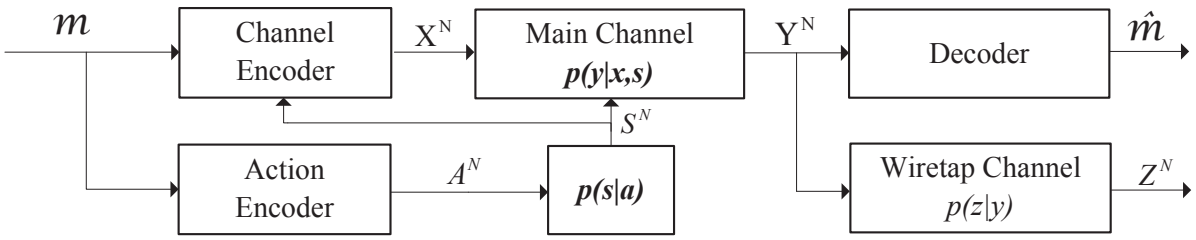
$$C_{dain} = \max_{p(u,v,q,a,x,s)} \min\{I(U;Y) - I(U;S|A), I(U;Y) - I(V;Z|Q)\} \tag{27}$$

Substituting $R_1 = 0$ into $\mathcal{R}_{on}$, we get a new upper bound on the secrecy capacity of the model in Figure 3 with noncausal states. This new upper bound is:

$$\begin{aligned} C'_{un} &= \max_{(R,R_e=R)\in\mathcal{R}_{nd}} R \\ &= \max_{p(u,v,q,a,x,s)} \min\{I(U;Y) - I(U;S|A), I(U;Y) - I(V;Z|Q) - I(U;S|V)\} &(28)\end{aligned}$$

where $\mathcal{R}_{nd}$ is the capacity-equivocation region of the model with noncausal states. Note that the difference between Equations (28) and (27) is the extra term, $I(U;S|V)$. The cause of the emergence of this term is stated in detail at the end of Appendix A. Then, we give the following corollary.

Figure 3. The wiretap channel with action-dependent states [18].



Corollary 4 *For the wiretap channel with noncausal action-dependent states shown in Figure 3, the new upper bound on the secrecy capacity* $C'_{un} \leq C_{dain}$.

Proof: *From the two Formulas, (27) and (28), we can get the difference between* C'_{un} *and* C_{dain} *as:*

$$\begin{aligned}\Lambda_1 &= C_{dain} - C'_{un} \\ &= \max\min\{I(U;Y) - I(U;S|A), I(U;Y) - I(V;Z|Q)\} \\ &\quad - \max\min\{I(U;Y) - I(U;S|A), I(U;Y) - I(V;Z|Q) - I(U;S|V)\}.\end{aligned}$$

It is easy to see that $\Lambda_1 \geq 0$, i.e., *our new upper bound* $C'_{un} \leq C_{dain}$. *Note that it is always desired to find a smaller upper bound to approach the secrecy capacity.*

Similarly, substituting $R_1 = 0$ into $\mathcal{R}_{oc}$, we can get an upper bound on the secrecy capacity of the model in Figure 3 with causal states, which is:

$$\begin{aligned}C'_{uc} &= \max_{(R,R_e=R)\in\mathcal{R}_{cd}} R \\ &= \max_{p(x|s,u)p(u,v,q,a,s)} I(U;Y) - I(V;Z|Q)\end{aligned} \tag{29}$$

where $\mathcal{R}_{cd}$ is the capacity-equivocation region of the model with causal states. C'_{uc} coincides with the upper bound on the secrecy capacity of the model in [18] with causal states.

Then, we study a special channel model where the "action" is removed, *i.e.*, the wiretap channel with noncausal channel state information. In this model, the channel state is generated by nature. This model is a special case of the wiretap channel with information embedding on actions by eliminating the action encoder and the mapping, f. It is also a special case of the model in Figure 3 without action. Setting the random variable, A, in Equation (28) to be a constant, we get a new outer bound of the wiretap channel with noncausal channel state information as:

$$C''_{un} = \max_{p(u,v,q,x,s)} \min\{I(U;Y) - I(U;S), I(U;Y) - I(V;Z|Q) - I(U;S|V)\}. \tag{30}$$

The outer bound in [20] was derived as $C'_{dain} = \max_{p(u,v,q,x,s)} \min\{I(U;Y) - I(U;S), I(U;Y) - I(V;Z|Q)\}$. Comparing the two bounds, we see that $C''_{un} \leq C'_{dain}$. This is stated in the following corollary.

Corollary 5 *For the wiretap channel with noncausal channel state information [20], the new upper bound on the secrecy capacity* $C''_{un} \leq C'_{dain}$.

In addition, in the case of no actions, our model turns into a special case that was also studied in [21]. The comparison between our results and those in [21] for this case is stated as follows.

- The achievable rate-equivocation region of the special case obtained from [21] is contained in that obtained from our results.
- We also provide an outer bound for this special case.

[21] obtained an achievable rate region for the broadcast wiretap channel with the asymmetric side information, which was:

$$\begin{aligned}
&\mathcal{R}_I = \{(R_1, R_2) \\
&R_1 \leq I(U_1; Y_1, S_1) - \max(I(U_1; Z), I(U_1; S_1, S_2)) && (31)\\
&R_2 \leq I(U_2; Y_2, S_2) - \max(I(U_2; Z), I(U_2; S_1, S_2)) && (32)\\
&R_1 + R_2 \leq I(U_1; Y_1, S_1) + I(U_2; Y_2, S_2) - I(U_1; U_2) \\
&\qquad - \max(I(U_1, U_2; Z), I(U_1, U_2; S_1, S_2))\} && (33)
\end{aligned}$$

We first give an equivalent expression of the achievable rate region $\mathcal{R}_I$ as follows.

$$\begin{aligned}
&\mathcal{R}'_I = \{(R_1, R_2, R_{e1}, R_{e2}) \\
&R_1 \leq I(U_1; Y_1, S_1) - \max(I(U_1; Z), I(U_1; S_1, S_2)) && (34)\\
&R_2 \leq I(U_2; Y_2, S_2) - \max(I(U_2; Z), I(U_2; S_1, S_2)) && (35)\\
&R_1 + R_2 \leq I(U_1; Y_1, S_1) + I(U_2; Y_2, S_2) - I(U_1; U_2) \\
&\qquad - \max(I(U_1, U_2; Z), I(U_1, U_2; S_1, S_2)) && (36)\\
&R_{e1} \leq R_1 && (37)\\
&R_{e2} \leq R_2 && (38)\\
&R_{e1} + R_{e2} \leq R_1 + R_2\} && (39)
\end{aligned}$$

where R_{e1} and R_{e2} are the equivocation rates of the messages, m_1 and m_2, respectively. We can easily prove that $\mathcal{R}'_I$ is equivalent to $\mathcal{R}_I$ via replacing Formulas (9), (10) and (11) in [21] accordingly by:

$$\begin{aligned}
\lim_{N\to\infty} \frac{H(m_1|Z^N)}{N} &\geq R_{e1} - \epsilon \\
\lim_{N\to\infty} \frac{H(m_2|Z^N)}{N} &\geq R_{e2} - \epsilon \\
\lim_{N\to\infty} \frac{H(m_1, m_2|Z^N)}{N} &\geq R_{e1} + R_{e2} - \epsilon
\end{aligned}$$

Since information is embedded on the actions, the information embedding rate $R_1 = 0$ when no actions are imposed in our model. At the same time, the specific decoder (Decoder 1) is no longer

needed. For this special case, we get its achievable rate-equivocation region from our results (by setting $R_1 = 0$ and $A = const$ in $\mathcal{R}_{in}$) as:

$$\begin{aligned}
&\mathcal{R}_{special} = \{(R_2, R_e) \\
&R_2 \leq I(U;Y) - I(U;S) && (40)\\
&R_e \leq R_2 && (41)\\
&R_e \leq I(U;Y) - \max\{I(U;Z), I(U;S)\}\} && (42)
\end{aligned}$$

As stated in [21], by removing the receiver ($\mathcal{D}_1$) and the side information ($S_2 = \mathcal{C}$) in the model considered in [21], we also arrive at the special case. Removing R_1, R_{e1}, Y_1, U_1 and S_2 in $\mathcal{R}'_I$, one has an achievable rate-equivocation region as:

$$\begin{aligned}
&\mathcal{R}'_{special} = \{(R_2, R_{e2}) \\
&R_2 \leq I(U_2;Y_2) - \max(I(U_2;Z), I(U_2,S_1)) && (43)\\
&R_{e2} \leq R_2\} && (44)
\end{aligned}$$

For simplicity, we replace U_2, S_1, R_{e2} and Y_2 by U, S, R_e and Y, respectively. We then show that $\mathcal{R}'_{special} \subseteq \mathcal{R}_{special}$. For any rate pair $(R_2, R_e) \in \mathcal{R}'_{special}$, from Equation (43),

$$\begin{aligned}
R_2 &\leq I(U;Y) - \max(I(U;Z), I(U,S)) \\
&\leq I(U;Y) - I(U,S) && (45)
\end{aligned}$$

From Equations (43) and (44),

$$\begin{aligned}
R_e &\leq R_2 \\
&\leq I(U;Y) - \max(I(U;Z), I(U,S)) && (46)
\end{aligned}$$

Therefore, $(R_2, R_e) \in \mathcal{R}_{special}$. This verifies $\mathcal{R}'_{special} \subseteq \mathcal{R}_{special}$.

Moreover, we get an outer bound for the special case. The outer bound is:

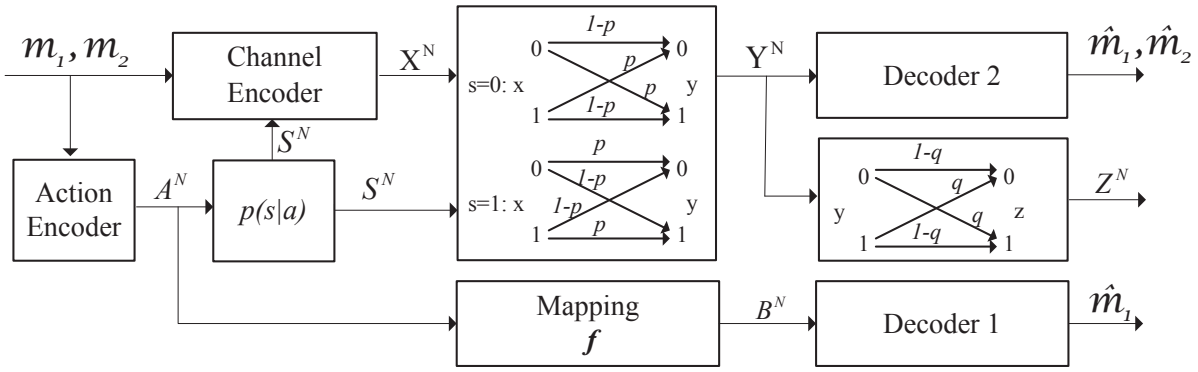
$$\begin{aligned}
&\mathcal{R}_{outer} = \{(R_2, R_e) \\
&R_2 \leq I(U;Y) - I(U;S) \\
&R_e \leq R_2 \\
&R_e \leq I(U;Y) - I(V;Z|Q) - I(U;S|V)\}
\end{aligned}$$

It can be directly gotten from $\mathcal{R}_{on}$ by setting $R_1 = 0$ and $A = const$. Note that [21] did not provide an outer bound.

3.2. A Binary Example

We give an example of a binary symmetric channel with causal channel states. The channel model is shown in Figure 4. Let the main channel be a binary symmetric channel (BSC). Its crossover probability is affected by the channel states. The wiretap channel is also assumed to be a BSC with crossover probability q. More precisely, define:

Figure 4. The binary symmetric channel with information embedding on actions.



$$p(y|x,s=i)=\begin{cases}(1-p)(1-i)+pi, & \text{if } y=x\\ (1-p)i+p(1-i), & \text{otherwise}\end{cases} \tag{47}$$

and:

$$p(z|y)=\begin{cases}1-q, & \text{if } z=y\\ q, & \text{otherwise}\end{cases} \tag{48}$$

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

It is assumed that the channel from the action to the channel states is a BSC with crossover probability equal to α, where $0\leq\alpha\leq 1$. In this example, the parameter, α, is fixed as 0.2 (the other value of α could also be assumed). Similar to the arguments in [1,18,19], the maximum values of $H(A|Z)$, $I(U;Y)$ and $I(U;Y)-I(U;Z)$ are achieved when $g_1:\mathcal{U}\to\mathcal{A}$ and $g_2:\mathcal{U}\times\mathcal{S}\to\mathcal{X}$ are deterministic mappings. We choose g_1 and g_2 as:

$$g_1(u=i)=i$$
$$g_2(u=i,s=j)=i+j \pmod 2$$

where $i,j\in\{0,1\}$. Let $B\sim$ Bernoulli(β), where $0\leq\beta\leq 1$. Let the function, f, be a one-to-one mapping for simplicity. Here, we set:

$$f:\begin{cases}0\to 0\\ 1\to 1\end{cases}$$

From the above conditions, we see that the random variables, B, A and U, share the same distribution. Then, the joint distribution $p(a,b,s,u,x,y,z)=p(z|y)p(y|x,s)p(x|u,s)p(s|a)p(a|u)p(u)1_{\{b=f(a)\}}$ can be calculated. The joint distributions $p(u,y)=\sum_{a,b,s,x,z}p(a,b,s,u,x,y,z)$ and $p(u,z)=\sum_{a,b,s,x,y}p(a,b,s,u,x,y,z)$ can also be obtained. By some mathematical calculation and Theorem 3, we can get the maximum sum secrecy rate of the example with causal channel states for given p,q as:

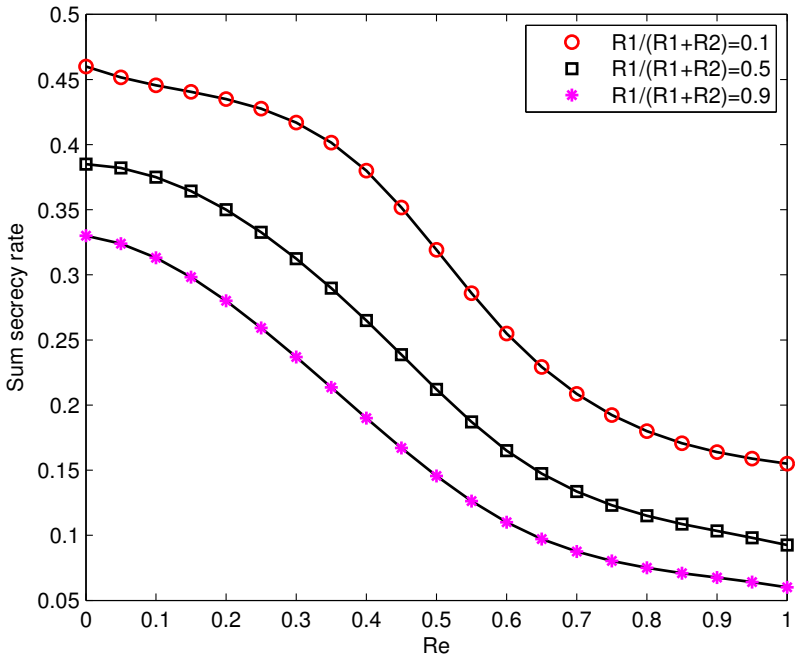
$$R_1+R_2=\max_{0\leq\beta\leq 1}\{\frac{5}{2}[h(q*(\beta*p))-h(\beta*p)]-\frac{3}{2}[h(p*q)-h(p)]\} \tag{49}$$

under the constraint $H(B)=h(\beta)\geq R_1$ and the secrecy constraint:

$$R_e\leq\min\{\max_{0\leq\beta\leq 1}\{h(p*q)-h(p),\frac{5}{2}[h(q*(\beta*p))-h(\beta*p)]-\frac{3}{2}[h(p*(0.2*q))-h(0.2*q)]\}$$
$$\max_{0\leq\beta\leq 1}\{h(\beta)-1+h(q*(\beta*p)),1-\frac{5}{2}[h(q*(\beta*p))-h(\beta*p)]\}\} \tag{50}$$

where $p*q = p+q-2pq$ and $h(p)$ is the binary entropy function, *i.e.*, $h(p) = -p\log p-(1-p)\log(1-p)$.

Figure 5. The tradeoff between the sum secrecy rate and the information embedding rate under the secrecy constraint.



The tradeoff between the sum secrecy rate and the information embedding rate under the secrecy constraint is shown in Figure 5. It can be seen that the sum secrecy rate is reduced when the equivocation rate, R_e, increases. In practical communication systems involving security, we always desire a bigger secure transmission rate when the extent of information leakage is at a reasonable level. Moreover, it can be seen that when the information embedding rate, R_1, goes up, the sum secrecy rate also decreases. This tells us that the communication requirements of Decoder 1 have a negative impact on improving the secrecy transmission rate of the given communication link.

4. Conclusions

This paper studies the wiretap channel with information embedding on actions. In this extended setup, the confidential message needs to be decoded only by the receiver and kept secret from the wiretapper as much as possible. Meanwhile, a specific decoder is introduced in the model to observe a function of the actions and wishes to decode part of the transmitted message. Our channel model is actually an extension of Ahmadi's channel with information embedding [11] by considering the secrecy constraint. We get the inner and outer bounds on the capacity-equivocation region of such a channel with noncausal (and causal) channel states. The corresponding lower and upper bounds on the sum secrecy capacity are also obtained. Besides, through a special case, we get a new upper bound on the secrecy capacity of the wiretap channel with action-dependent states and show that it is tighter than the upper bound obtained in [18]. We also discuss the tradeoff between the sum secrecy rate and the information embedding rate under the secrecy constraint.

Some potential directions that are worthy of being explored are listed as follows.

- In practical application, the wiretapper may also wish to eavesdrop on the embedded information. In the example of communication networks, the information embedded in the packet for the next router may also be of interest to the eavesdropper. Our current setting does not consider the confidentiality of m_1 and m_2 separately, so this problem will be further explored.
- Only inner and outer bounds on the capacity-equivocation region are obtained at present. We can try to find some special cases where the two bounds match. Moreover, if there exists a channel between A^N and B^N instead of a function, what will the capacity-equivocation region be?
- Adaptive action means that the action sequence is generated by the message and the previous channel states, *i.e.*, $a_i(m, s^{i-1})$. 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 [10], 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 sum secrecy capacity of our channel model under the secrecy constraint.

Appendix

A. Proof of Theorems 1 and 2

In this section, Theorems 1 and 2 are proven. To prove Theorem 1, the methods in [18] are utilized, and we present a coding scheme for the model in Figure 2 with noncausal action-dependent states in Subsection A.1. The proof of Theorem 2 is given in Subsection A.2.

A.1. Proof of Theorem 1

We need to prove that any rate-equivocation triple $(R_1, R_2, R_e) \in \mathcal{R}_{in}$ is achievable. Similar to [18], two cases are considered. Since the channel states are noncausally known to the channel encoder, Gel'fand and Pinsker's coding technique [22] will be used in the encoding process.

A.1.1. $H(A|Z) \geq I(U;Y) - \max\{I(U;Z), I(U;S|A)\}$

In this case, we need to prove that any rate-equivocation triple (R_1, R_2, R_e) satisfying the following constraints are achievable.

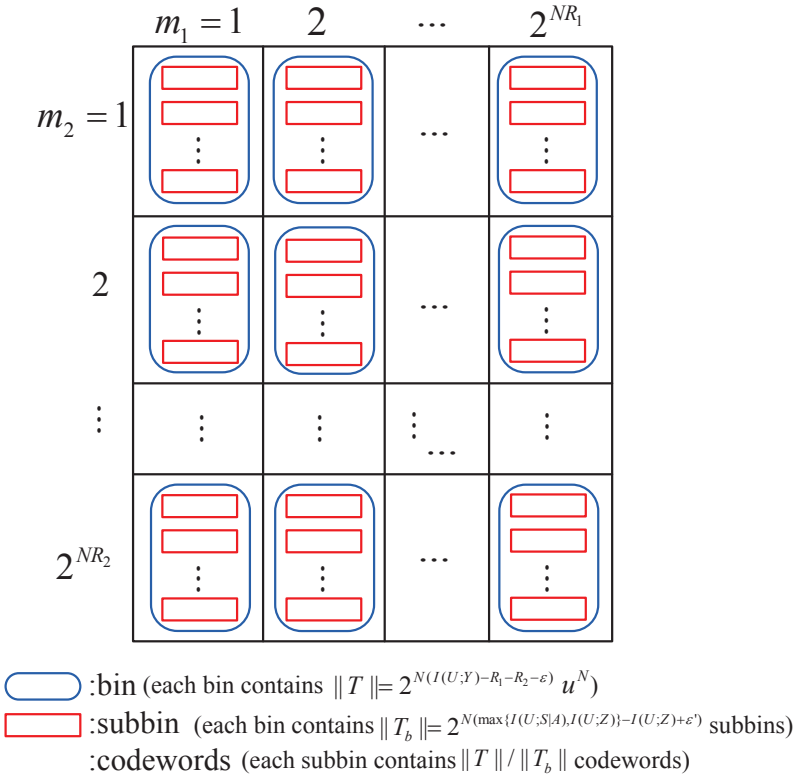
$$\begin{aligned} R_1 &\leq H(B) \\ R_1 + R_2 &\leq I(U;Y) - I(U;S|A) \\ R_e &\leq R_1 + R_2 \\ R_e &\leq I(U;Y) - \max\{I(U;Z), I(U;S|A)\} \end{aligned}$$

It is sufficient to show that the rate triples $(R_1, R_2, R_e = I(U;Y) - \max\{I(U;Z), I(U;S|A)\})$ are achievable. The coding scheme includes codebook generation, encoding and decoding. Then we give the equivocation analysis.

Codebook generation and encoding: Let $R_1 = H(B) - \tau_1$ and $R_1 + R_2 = I(U;Y) - I(U;S|A) - \tau_2$, where τ_1, τ_2 are fixed positive numbers. Since $R_e \leq R_1 + R_2$, it is easy to get $\tau_2 \leq$

$\max\{I(U;S|A), I(U;Z)\} - I(U;S|A)$. For each $m_1 \in \{1, 2, ..., 2^{NR_1}\}$, an independent and identically distributed (i.i.d) codeword, $b^N(m_1)$, is generated according to $p(b^N) = \prod_{i=1}^{N} p(b_i)$. Then, 2^{NR_2} action sequences $a^N(m_1, m_2)$ are i.i.d generated for each $b^N(m_1)$ according to $p(a^N(m_1, m_2)|b^N(m_1)) = \prod_{i=1}^{N} p(a_i|b_i)$, where $m_2 \in \{1, 2, ..., 2^{NR_2}\}$. For each $a^N(m_1, m_2)$, we generate $\|\mathcal{T}\| = 2^{N(I(U;Y)-R_1-R_2-\epsilon)}$ i.i.d codewords $u^N(m_1, m_2, t_b, t_u)$ according to $p(u^N(m_1, m_2, t_b, t_u)|a^N(m_1, m_2)) = \prod_{i=1}^{N} p(u_i|a_i)$. These codewords are put into $\|\mathcal{T}_b\| = 2^{N(\max\{I(U;S|A),I(U;Z)\}-I(U;Z)+\epsilon')}$ bins, such that each bin contains $\|\mathcal{T}\|/\|\mathcal{T}_b\|$ codewords. Note that t_b, t_u are the indexes of the bin and codeword, respectively. The codebook structure is shown in Figure A1. To send the message (m_1, m_2) with the action sequence, $a^N(m_1, m_2)$, and corresponding state sequence s^N, the encoder chooses a $u^N(m_1, m_2, t_b, t_u)$ from the $\|\mathcal{T}\|$ sequences, such that $(u^N(m_1, m_2, t_b, t_u), a^N(m_1, m_2), s^N) \in T^N_{US|A}$. If no such sequence exists, it picks $(t_b, t_u) = (1, 1)$. Then, the input sequence of the channel is generated by $p(x^N|u^N, s^N) = \prod_{i=1}^{N} p(x_i|u_i, s_i)$.

Figure A1. Codebook structure.



Decoding and error probability analysis: Decoder 1 can decode the message, m_1, correctly, since $R_1 \leq H(B)$. For the receiver, he tries to find a unique sequence $u^N(\hat{m}_1, \hat{m}_2, \hat{t}_b, \hat{t}_u)$, such that $(u^N(\hat{m}_1, \hat{m}_2, \hat{t}_b, \hat{t}_u), a^N(\hat{m}_1, \hat{m}_2), y^N) \in T^N_{UAY}$. It is easy to show the decoding error probabilities $P_{e1} \leq \epsilon$ and $P_{e2} \leq \epsilon$ by similar arguments in [11,12]. We mainly focus on the analysis of equivocation.

Equivocation analysis:

$$\begin{aligned}
H(M_1, M_2|Z^N) &= H(M_1, M_2, Z^N) - H(Z^N) \\
&= H(M_1, M_2, Z^N, U^N) - H(U^N|M_1, M_2, Z^N) - H(Z^N) \\
&= H(M_1, M_2, U^N) + H(Z^N|M_1, M_2, U^N) - H(U^N|M_1, M_2, Z^N) - H(Z^N) \\
&\geq H(U^N) + H(Z^N|U^N) - H(U^N|M_1, M_2, Z^N) - H(Z^N) \qquad (51) \\
&\geq H(U^N) - I(U^N; Z^N) - H(T_b, U^N|M_1, M_2, Z^N) \\
&= H(U^N) - I(U^N; Z^N) - H(T_b|M_1, M_2, Z^N) - H(U^N|M_1, M_2, Z^N, T_b) \\
&\geq I(U^N; Y^N) - I(U^N; Z^N) - H(T_b) - H(U^N|M_1, M_2, Z^N, T_b) \\
&\geq NI(U; Y) - NI(U; Z) - H(T_b) - H(U^N|M_1, M_2, Z^N, T_b), \qquad (52)
\end{aligned}$$

where Equation (51) is from the Markov chain $(M_1, M_2) \to U^N \to Z^N$, and Equation (52) is from that the codewords, u^N, are i.i.d and the channels are discrete memoryless.

Next, we bound $H(T_b)$ and $H(U^N|M_1, M_2, Z^N, T_b)$. Since $\|\mathcal{T}_b\| = 2^{N(\max\{I(U;S|A), I(U;Z)\} - I(U;Z) + \epsilon')}$, we have $H(T_b) \leq log\|\mathcal{T}_b\| = N(\max\{I(U; S|A), I(U; Z)\} - I(U; Z) + \epsilon')$.

The explanation for bounding $\frac{1}{N}H(U^N|M_1, M_2, Z^N, T_b)$ is presented as follows. We first show that, given M_1,M_2 and T_b, the probability of error for Z^N to decode U^N satisfies $P_e \leq \nu$. Here, ν is small for sufficiently large N. Given the knowledge of M_1,M_2 and T_b, the total number of possible codewords of U^N is:

$$\begin{aligned}
\frac{\|\mathcal{T}\|}{\|\mathcal{T}_b\|} &= \frac{2^{N(I(U;Y) - R_1 - R_2 - \epsilon)}}{2^{N(\max\{I(U;S|A), I(U;Z)\} - I(U;Z) + \epsilon')}} \\
&= \frac{2^{N(I(U;S|A) + \tau_2 - \epsilon)}}{2^{N(\max\{I(U;S|A), I(U;Z)\} - I(U;Z) + \epsilon')}} \\
&\leq \frac{2^{N(\max\{I(U;S|A), I(U;Z)\} - \epsilon)}}{2^{N(\max\{I(U;S|A), I(U;Z)\} - I(U;Z) + \epsilon')}} \qquad (53) \\
&= 2^{N(I(U;Z) - \epsilon - \epsilon')} \\
&\leq 2^{NI(U;Z)} \qquad (54)
\end{aligned}$$

where Equation (53) follows from $\tau_2 \leq \max\{I(U; S|A), I(U; Z)\} - I(U; S|A)$. Based on Equation (54), we can easily show that a unique codeword $u^N(m_1, m_2, t_b, t_u)$ exists, such that $(u^N(m_1, m_2, t_b, t_u), z^N) \in T^N_{UZ}$ with high probability. This indicates that the probability of error for Z^N to decode U^N satisfies $P_e \leq \nu$. Therefore, by Fano's inequality, we obtain:

$$\frac{1}{N}H(U^N|M_1, M_2, Z^N, T_b) \leq \frac{1}{N}(1 + P_e log(\frac{\|\mathcal{T}\|}{\|\mathcal{T}_b\|})) \leq \nu' \qquad (55)$$

where ν' is small for sufficiently large N.

Substituting these two results into Equation (52) and utilizing Equation (3), we finish the proof of $\lim_{N\to\infty} \triangle \geq R_e$ for the model in Figure 2 with noncausal channel states.

A.1.2. $H(A|Z) \leq I(U;Y) - \max\{I(U;Z), I(U;S|A)\}$

In this case, we need to prove that any rate-equivocation triple (R_1, R_2, R_e) satisfying the following constraints are achievable.

$$\begin{aligned} R_1 &\leq H(B) \\ R_1 + R_2 &\leq I(U;Y) - I(U;S|A) \\ R_e &\leq R_1 + R_2 \\ R_e &\leq H(A|Z) \end{aligned}$$

It is sufficient to show that the rate triples $(R_1, R_2, R_e = H(A|Z))$ are achievable. The coding scheme is as follows.

Codebook generation and encoding: Let $R_1 = H(B) - \tau_1'$ and $R_1 + R_2 = I(U;Y) - I(U;S|A) - \tau_2'$, where τ_1', τ_2' are fixed positive numbers. For each $m_1 \in \{1, 2, ..., 2^{NR_1}\}$, an independent and identically distributed (i.i.d) codeword, $b^N(m_1)$, is generated according to $p(b^N) = \prod_{i=1}^{N} p(b_i)$. Then, 2^{NR_2} action sequences $a^N(m_1, m_2)$ are i.i.d generated for each $b^N(m_1)$ according to $p(a^N(m_1, m_2)|b^N(m_1)) = \prod_{i=1}^{N} p(a_i|b_i)$, where $m_2 \in \{1, 2, ..., 2^{NR_2}\}$. For each $a^N(m_1, m_2)$, we generate $\|\mathcal{T}\| = 2^{N(I(U;Y)-R_1-R_2-\epsilon)}$ i.i.d codewords $u^N(m_1, m_2, t_u)$ according to $p(u^N(m_1, m_2, t_u)|a^N(m_1, m_2)) = \prod_{i=1}^{N} p(u_i|a_i)$. To send the message (m_1, m_2) with the action sequence, $a^N(m_1, m_2)$, and corresponding state sequence s^N, the encoder chooses a $u^N(m_1, m_2, t_u)$ from the $\|\mathcal{T}\|$ sequences, such that $(u^N(m_1, m_2, t_u), a^N(m_1, m_2), s^N) \in T^N_{US|A}$. If no such sequence exists, it picks $t_u = 1$. Then, the input sequence of the channel is generated by $p(x^N|u^N, s^N) = \prod_{i=1}^{N} p(x_i|u_i, s_i)$.

Decoding and error probability analysis: Decoder 1 can decode the message, m_1, correctly, since $R_1 \leq H(B)$. For the receiver, he tries to find a unique sequence $u^N(\hat{\hat{m}}_1, \hat{\hat{m}}_2, \hat{\hat{t}}_u)$, such that $(u^N(\hat{\hat{m}}_1, \hat{\hat{m}}_2, \hat{\hat{t}}_u), a^N(\hat{\hat{m}}_1, \hat{\hat{m}}_2), y^N) \in T^N_{UAY}$. It is easy to show the decoding error probabilities $P_{e1} \leq \epsilon$ and $P_{e2} \leq \epsilon$ by similar arguments in [11,12].

Equivocation analysis: We need to prove $\lim_{N\to\infty} \triangle = \lim_{N\to\infty} \frac{H(M_1, M_2|Z^N)}{N} \geq R_e$. The methods in [18] are utilized.

$$\begin{aligned} \lim_{N\to\infty} \frac{H(M_1, M_2|Z^N)}{N} &= \lim_{N\to\infty} \frac{H(A^N(M_1, M_2)|Z^N)}{N} && (56) \\ &= \lim_{N\to\infty} \frac{NH(A|Z)}{N} && (57) \\ &= H(A|Z) \\ &\geq R_e \end{aligned}$$

where Equation (56) is from that A^N is a function of (M_1, M_2), and Equation (57) is from that the sequences A^N and X^N are i.i.d generated and the channels are discrete memoryless.

We complete the proof of Theorem 1.

A.2. Proof of Theorem 2

In this subsection, we prove that all achievable rate triples (R_1, R_2, R_e) for the model in Figure 2 with noncausal channel states are contained in $\mathcal{R}_{on}$.

To prove condition in Equation (11), we consider:

$$\begin{aligned} R_1 &= \lim_{N\to\infty} \frac{\log \|\mathcal{M}_1\|}{N} \\ &= \lim_{N\to\infty} \frac{H(M_1)}{N} \\ &= \lim_{N\to\infty} \frac{1}{N}[I(M_1; B^N) + H(M_1|B^N)] \\ &\leq \lim_{N\to\infty} \frac{1}{N}[I(M_1; B^N) + \delta(P_{e1})] \qquad (58) \\ &\leq \lim_{N\to\infty} \frac{1}{N}[H(B^N) + \delta(P_{e1})] \\ &= \lim_{N\to\infty} \frac{1}{N}[\sum_{i=1}^{N} H(B_i|B^{i-1}) + \delta(P_{e1})] \\ &\leq \lim_{N\to\infty} \frac{1}{N}[\sum_{i=1}^{N} H(B_i) + \delta(P_{e1})] \qquad (59) \end{aligned}$$

where Equation (58) is based on Fano's inequality.

To prove the condition in Equation (12), we consider:

$$\begin{aligned} R_1 + R_2 &= \lim_{N\to\infty} \frac{\log(\|\mathcal{M}_1\| \cdot \|\mathcal{M}_2\|)}{N} \\ &= \lim_{N\to\infty} \frac{H(M_1, M_2)}{N} \\ &= \lim_{N\to\infty} \frac{1}{N}[I(M_1, M_2; Y^N) + H(M_1, M_2|Y^N)] \\ &\leq \lim_{N\to\infty} \frac{1}{N}[I(M_1, M_2; Y^N) + \delta(P_{e2})] \qquad (60) \end{aligned}$$

where Equation (60) is based on Fano's inequality. Then, the mutual information $I(M_1, M_2; Y^N)$ in Equation (60) is calculated as follows.

$$\begin{aligned} I(M_1, M_2; Y^N) &= I(M_1, M_2; Y^N) - I(M_1, M_2; S^N|A^N) \qquad (61) \\ &= \sum_{i=1}^{N} [I(M_1, M_2; Y_i|Y^{i-1}) - I(M_1, M_2; S_i|S_{i+1}^N, A^N)] \\ &= \sum_{i=1}^{N} [I(M_1, M_2, S_{i+1}^N, A^N; Y_i|Y^{i-1}) - I(S_{i+1}^N, A^N; Y_i|M_1, M_2, Y^{i-1}) \\ &\quad - I(M_1, M_2, Y^{i-1}; S_i|S_{i+1}^N, A^N) + I(Y^{i-1}; S_i|M_1, M_2, S_{i+1}^N, A^N)] \\ &= \sum_{i=1}^{N} [I(M_1, M_2, S_{i+1}^N, A^N; Y_i|Y^{i-1}) - I(M_1, M_2, Y^{i-1}; S_i|S_{i+1}^N, A^N)] \qquad (62) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N, Y^{i-1}; Y_i) - I(M_1, M_2, Y^{i-1}; S_i|S_{i+1}^N, A^N)] \\
&= \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N, Y^{i-1}; Y_i) - I(M_1, M_2, Y^{i-1}, S_{i+1}^N, A^N; S_i|A_i) \\
&\quad + I(S_{i+1}^N, A^N; S_i|A_i)] \\
&= \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N, Y^{i-1}; Y_i) - I(M_1, M_2, Y^{i-1}, S_{i+1}^N, A^N; S_i|A_i)] \qquad (63) \\
&= \sum_{i=1}^{N}[I(U_i; Y_i) - I(U_i; S_i|A_i)] \qquad (64)
\end{aligned}
$$

In the above deduction, Equation (61) is from the Markov chain $(M_1, M_2) \rightarrow A^N \rightarrow S^N$. The Equation (62) is from the $\sum_{i=1}^{N} I(S_{i+1}^N, A^N; Y_i|M_1, M_2, Y^{i-1}) = \sum_{i=1}^{N} I(Y^{i-1}; S_i|M_1, M_2, S_{i+1}^N, A^N)$, which can be derived similarly according to [1] and [18]. The Equation (63) is from the Markov chain $S_i \rightarrow A_i \rightarrow (S_{i+1}^N, A^{i-1}, A_{i+1}^N)$. The Equation (64) is from defining $U_i = (M_1, M_2, Y^{i-1}, S_{i+1}^N, A^N)$.

The condition in Equation (13) is proven as follows.

$$
\begin{aligned}
R_e &\leq \lim_{N\to\infty} \Delta \\
&= \lim_{N\to\infty} \frac{H(M_1, M_2|Z^N)}{N} \\
&\leq \lim_{N\to\infty} \frac{H(M_1, M_2)}{N} = R_1 + R_2 \qquad (65)
\end{aligned}
$$

The condition in Equation (14) is proven as follows.

$$
\begin{aligned}
H(M_1, M_2|Z^N) &= H(M_1, M_2|Z^N) - H(M_1, M_2|Y^N) + H(M_1, M_2|Y^N) \\
&= H(M_1, M_2|Z^N) - H(M_1, M_2) + H(M_1, M_2) - H(M_1, M_2|Y^N) + H(M_1, M_2|Y^N) \\
&= I(M_1, M_2; Y^N) - I(M_1, M_2; Z^N) + H(M_1, M_2|Y^N) \\
&\leq I(M_1, M_2; Y^N) - I(M_1, M_2; Z^N) + \delta(P_{e2}) \qquad (66)
\end{aligned}
$$

From Equation (62), we have:

$$
I(M_1, M_2; Y^N) = \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N; Y_i|Y^{i-1}) - I(M_1, M_2, Y^{i-1}; S_i|S_{i+1}^N, A^N)] \qquad (67)
$$

Similarly, we can get:

$$
I(M_1, M_2; Z^N) = \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N; Z_i|Z^{i-1}) - I(M_1, M_2, Z^{i-1}; S_i|S_{i+1}^N, A^N)] \qquad (68)
$$

Substitute Equations (67) and (68) into Equation (66),

$$\begin{aligned}
H(M_1, M_2|Z^N) \leq & \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N; Y_i|Y^{i-1}) - I(M_1, M_2, Y^{i-1}; S_i|S_{i+1}^N, A^N) \\
& - I(M_1, M_2, S_{i+1}^N, A^N; Z_i|Z^{i-1}) + I(M_1, M_2, Z^{i-1}; S_i|S_{i+1}^N, A^N)] + \delta(P_{e2}) \\
= & \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N; Y_i|Y^{i-1}) - I(M_1, M_2, Z^{i-1}, Y^{i-1}; S_i|S_{i+1}^N, A^N) \\
& - I(M_1, M_2, S_{i+1}^N, A^N; Z_i|Z^{i-1}) + I(M_1, M_2, Z^{i-1}; S_i|S_{i+1}^N, A^N)] + \delta(P_{e2}) \quad (69) \\
= & \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N; Y_i|Y^{i-1}) - I(M_1, M_2, S_{i+1}^N, A^N; Z_i|Z^{i-1}) \\
& - I(Y^{i-1}; S_i|M_1, M_2, Z^{i-1}, S_{i+1}^N, A^N)] + \delta(P_{e2}) \\
\leq & \sum_{i=1}^{N}[I(M_1, M_2, S_{i+1}^N, A^N, Y^{i-1}; Y_i) - I(M_1, M_2, S_{i+1}^N, A^N, Z^{i-1}; Z_i|Z^{i-1}) \\
& - I(M_1, M_2, S_{i+1}^N, A^N, Y^{i-1}; S_i|M_1, M_2, Z^{i-1}, S_{i+1}^N, A^N)] + \delta(P_{e2}) \quad (70) \\
= & I(U_i; Y_i) - I(V_i; Z_i|Q_i) - I(U_i; S_i|V_i) + \delta(P_{e2}) \quad (71)
\end{aligned}$$

where Equation (69) is from the Markov chain $S_i \rightarrow (S_{i+1}^N, A^N, M_1, M_2, Y^{i-1}) \rightarrow Z^{i-1}$ and Equation (71) is from defining $U_i = (M_1, M_2, S_{i+1}^N, A^N, Y^{i-1})$, $V_i = (M_1, M_2, S_{i+1}^N, A^N, Z^{i-1})$ and $Q_i = Z^{i-1}$.

To serve the single-letter characterization, let us introduce a time-sharing random variable, J, independent of all other random variables and uniformly distributed over $\{1, 2, ..., N\}$. Set:

$$\begin{aligned}
& U = (U_J, J), V = (V_J, J), Q = (Q_J, J) \\
& A = A_J, B = B_J, S = S_J, X = X_J, Y = Y_J, Z = Z_J
\end{aligned}$$

Then, substituting the above new random variables into Equations (59), (64) and (71), the conditions in Equations (11), (12) and (14) are verified. From the definition of the auxiliary random variables, the Markov chain $Q \rightarrow V \rightarrow U \rightarrow Y \rightarrow Z$ is easy to be verified. We complete the proof of Theorem 2.

We note that Dai *et al.* [18] got an upper bound on R_e for the model in Figure 3 as $I(U; Y) - I(V; Z|Q)$. Comparing it with our bound on R_e, we notice that an extra term, $I(U; S|V)$, is contained in our bound. The difference between our proof and the proof in [18] about the upper bounds is stated in detail as follows. We both concentrate on zooming $H(M_1, M_2|Z^N)$. In [18], $H(M|Z^N)$ was considered, since no embedding information was imposed. From Equation (66),

$$H(M_1, M_2|Z^N) \leq I(M_1, M_2; Y^N) - I(M_1, M_2; Z^N) + \delta(P_{e2}) \quad (72)$$

The two terms in Equation (72) are calculated *independently* in [18] as:

$$I(M_1, M_2; Y^N) \leq \sum_{i=1}^{N}[I(U_i; Y_i) - I(U_i; S_i|A_i)]$$

and:

$$I(M_1, M_2; Z^N) \geq \sum_{i=1}^{N} [I(V_i; Z_i|Q_i) - I(U_i; S_i|A_i)]$$

respectively. Then, the term, $I(U_i; S_i|A_i)$, was offset in [18] by subtraction in order to obtain their upper bound. However, the weak aspect of calculating the two terms independently is that the interrelation between the two terms are missed.

In our proof, we focus on calculating $I(M_1, M_2; Y^N) - I(M_1, M_2; Z^N)$. The result is shown in Equation (71). The main difference in the proof steps between our work and [18] is Equations (69) and (70). In the above derivation, we can see that the extra term, $I(U_i; S_i|V_i)$, is originated from $I(M_1, M_2, Y^{i-1}; S_i|S_{i+1}^N, A^N) - I(M_1, M_2, Z^{i-1}; S_i|S_{i+1}^N, A^N)$.

B. Proof of Theorems 3 and 4

In this section, Theorems 3 and 4 are proven. To prove Theorem 3, the methods in [18] are utilized and a coding scheme for the model in Figure 2 with causal action-dependent states is provided in Subsection *B.1*. Subsection *B.2* gives the proof of the outer bound on the capacity-equivocation region.

B.1. Proof of Theorem 3

We need to prove that any achievable rate triple $(R_1, R_2, R_e) \in \mathcal{R}_{ic}$ is achievable. Two cases are considered.

B.1.1. $H(A|Z) \geq I(U;Y) - I(U;Z)$

In this case, we need to prove that any rate-equivocation triple (R_1, R_2, R_e) satisfying the following constraints are achievable.

$$\begin{aligned} R_1 &\leq H(B) \\ R_1 + R_2 &\leq I(U;Y) \\ R_e &\leq R_1 + R_2 \\ R_e &\leq I(U;Y) - I(U;Z) \end{aligned}$$

It is sufficient to show that the rate triples $(R_1, R_2, R_e = I(U;Y) - I(U;Z))$ are achievable. The coding scheme includes codebook generation, encoding and decoding. Then, we give the equivocation analysis.

Codebook generation and encoding: Let $R_1 = H(B) - \theta_1$ and $R_1 + R_2 = I(U;Y) - \theta_2$, where θ_1, θ_2 are fixed positive numbers. Since $R_e \leq R_1 + R_2$, it is easy to get $\theta_2 \leq I(U;Z)$. For each $m_1 \in \{1, 2, ..., 2^{NR_1}\}$, an independent and identically distributed (i.i.d) codeword, $b^N(m_1)$, is generated according to $p(b^N) = \prod_{i=1}^{N} p(b_i)$. Then, 2^{NR_2} action sequences $a^N(m_1, m_2)$ are i.i.d generated for each $b^N(m_1)$ according to $p(a^N(m_1, m_2)|b^N(m_1)) = \prod_{i=1}^{N} p(a_i|b_i)$, where $m_2 \in \{1, 2, ..., 2^{NR_2}\}$. For each

$a^N(m_1, m_2)$, we generate $\|\mathcal{T}_u\| = 2^{N(I(U;Y)-R_1-R_2-\epsilon)}$ i.i.d codewords $u^N(m_1, m_2, t_u)$ according to $p(u^N(m_1, m_2, t_u)|a^N(m_1, m_2)) = \prod_{i=1}^{N} p(u_i|a_i)$. Note that t_u is the index of codeword u^N. To send the message (m_1, m_2) with the action sequence, $a^N(m_1, m_2)$, and corresponding state sequence s^N, the encoder randomly chooses an index $t_u^* \in \{1, 2, ..., \|\mathcal{T}_u\|\}$. Then, the input sequence of the channel is generated by $p(x^N|u^N(m_1, m_2, t_u^*), s^N) = \prod_{i=1}^{N} p(x_i|u_i, s_i)$.

Decoding and error probability analysis: Decoder 1 can decode the message, m_1, correctly, since $R_1 \leq H(B)$. For the receiver, he tries to find a unique sequence $u^N(\hat{\hat{m}}_1, \hat{\hat{m}}_2, \hat{\hat{t}}_u)$, such that $(u^N(\hat{\hat{m}}_1, \hat{\hat{m}}_2, \hat{\hat{t}}_u), a^N(\hat{\hat{m}}_1, \hat{\hat{m}}_2), y^N) \in T^N_{UAY}$. It is easy to show the decoding error probabilities $P_{e1} \leq \epsilon$ and $P_{e2} \leq \epsilon$, and therefore, we omit the proof here. We mainly focus on the analysis of equivocation.

Equivocation analysis:

$$\begin{aligned}
H(M_1, M_2|Z^N) &= H(M_1, M_2, Z^N) - H(Z^N) \\
&= H(M_1, M_2, Z^N, U^N) - H(U^N|M_1, M_2, Z^N) - H(Z^N) \\
&= H(M_1, M_2, U^N) + H(Z^N|M_1, M_2, U^N) - H(U^N|M_1, M_2, Z^N) - H(Z^N) \\
&\geq H(U^N) + H(Z^N|U^N) - H(U^N|M_1, M_2, Z^N) - H(Z^N) \qquad (73) \\
&= H(U^N) - I(U^N; Z^N) - H(U^N|M_1, M_2, Z^N) \\
&\geq I(U^N; Y^N) - I(U^N; Z^N) - H(U^N|M_1, M_2, Z^N) \\
&\geq NI(U; Y) - NI(U; Z) - H(U^N|M_1, M_2, Z^N) \qquad (74)
\end{aligned}$$

where Equation (73) is from the Markov chain $(M_1, M_2) \to U^N \to Z^N$, and Equation (74) is from that the codewords, u^N, are i.i.d and the channels are discrete memoryless. The conditional entropy, $H(U^N|M_1, M_2, Z^N)$, is calculated as follows. Given the message (M_1, M_2), the number of U^N is $\|\mathcal{T}_u\| = 2^{N(I(U;Y)-R_1-R_2-\epsilon)} = 2^{N(\theta_2-\epsilon)} \leq 2^{N(I(U;Z)-\epsilon)}$. Therefore, $H(U^N|M_1, M_2, Z^N) \to 0$ as $N \to \infty$. Substituting this result into Equation (74) and utilizing Equation (3), we finish the proof of $\lim_{N\to\infty} \triangle \geq R_e$ for the model in Figure 2 with causal channel states.

B.1.2. $H(A|Z) \leq I(U;Y) - I(U;Z)$

In this case, we need to prove that any rate-equivocation triple (R_1, R_2, R_e) satisfying the following constraints are achievable.

$$\begin{aligned}
R_1 &\leq H(B) \\
R_1 + R_2 &\leq I(U;Y) \\
R_e &\leq R_1 + R_2 \\
R_e &\leq H(A|Z)
\end{aligned}$$

It is sufficient to show that $(R_1, R_2, R_e = H(A|Z))$ are achievable. The coding scheme is as follows.

Codebook generation and encoding: Let $R_1 = H(B) - \theta_1'$ and $R_1 + R_2 = I(U;Y) - \theta_2'$, where θ_1', θ_2' are fixed positive numbers. For each $m_1 \in \{1, 2, ..., 2^{NR_1}\}$, an independent and identically distributed (i.i.d) codeword, $b^N(m_1)$, is generated according to $p(b^N) = \prod_{i=1}^{N} p(b_i)$. Then, the 2^{NR_2} action sequences

$a^N(m_1, m_2)$ are i.i.d generated for each $b^N(m_1)$ according to $p(a^N(m_1, m_2)|b^N(m_1)) = \prod_{i=1}^{N} p(a_i|b_i)$, where $m_2 \in \{1, 2, ..., 2^{NR_2}\}$. For each $a^N(m_1, m_2)$, a corresponding codeword, $u^N(m_1, m_2)$, is generated according to $p(u^N(m_1, m_2)|a^N(m_1, m_2)) = \prod_{i=1}^{N} p(u_i|a_i)$. To send the message (m_1, m_2) with the action sequence, $a^N(m_1, m_2)$, and corresponding state sequence s^N, the encoder selects the codeword, $u^N(m_1, m_2)$. Then, the input sequence of the channel is generated by $p(x^N|u^N(m_1, m_2), s^N) = \prod_{i=1}^{N} p(x_i|u_i, s_i)$.

Decoding and error probability analysis: This step follows similarly from Case *A* in Section IV.

Equivocation analysis: We need to prove $\lim_{N\to\infty} \triangle = \lim_{N\to\infty} \frac{H(M_1,M_2|Z^N)}{N} \geq R_e$. The methods in [18] are utilized.

$$\lim_{N\to\infty} \frac{H(M_1, M_2|Z^N)}{N} = \lim_{N\to\infty} \frac{H(A^N(M_1, M_2)|Z^N)}{N} \tag{75}$$

$$= \lim_{N\to\infty} \frac{NH(A|Z)}{N} \tag{76}$$

$$= H(A|Z)$$

$$\geq R_e$$

where Equation (75) is from that A^N is a function of (M_1, M_2), and Equation (76) is from that the sequences, A^N and X^N, are i.i.d generated and that the channels are discrete memoryless.

We complete the proof of Theorem 3.

B.2. Proof of Theorem 4

In this subsection, we need to prove that all achievable rate triples (R_1, R_2, R_e) for the model in Figure 2 with causal channel states are contained in $\mathcal{R}_{oc}$.

The conditions in Equations (16) and (18) follow the same as those of Equations (59) and (65). Therefore, we show Equations (17) and (19) as follows.

To prove the condition in Equation (17), we consider:

$$\begin{aligned} R_1 + R_2 &= \lim_{N\to\infty} \frac{\log(\|\mathcal{M}_1\| \cdot \|\mathcal{M}_2\|)}{N} \\ &= \lim_{N\to\infty} \frac{H(M_1, M_2)}{N} \\ &= \lim_{N\to\infty} \frac{1}{N}[I(M_1, M_2; Y^N) + H(M_1, M_2|Y^N)] \\ &\leq \lim_{N\to\infty} \frac{1}{N}[I(M_1, M_2; Y^N) + \delta(P_{e2})] \end{aligned} \tag{77}$$

$$
\begin{aligned}
&= \lim_{N\to\infty} \frac{1}{N}[\sum_{i=1}^{N} I(M_1, M_2; Y_i|Y^{i-1}) + \delta(P_{e2})] \\
&\leq \lim_{N\to\infty} \frac{1}{N}[\sum_{i=1}^{N} I(M_1, M_2, Y^{i-1}, S^{i-1}; Y_i) + \delta(P_{e2})], \\
&\leq \lim_{N\to\infty} \frac{1}{N}[\sum_{i=1}^{N} I(U_i; Y_i) + \delta(P_{e2})] \qquad (78)
\end{aligned}
$$

where Equation (77) is based on Fano's inequality and Equation (78) is from defining $U_i = (M_1, M_2, Y^{i-1}, S^{i-1})$.

Before proving the condition in Equation (19), we consider:

$$
\begin{aligned}
I(M_1, M_2; Z^N) &= \sum_{i=1}^{N} I(M_1, M_2; Z_i|Z^{i-1}) \\
&= \sum_{i=1}^{N} I(M_1, M_2, Z^{i-1}; Z_i|Z^{i-1}). \\
&= \sum_{i=1}^{N} I(V_i; Z_i|Q_i) \qquad (79)
\end{aligned}
$$

where Equation (79) is from defining $V_i = (M_1, M_2, Z^{i-1})$ and $Q_i = Z^{i-1}$.

Then, utilizing Equations (78) and (79),

$$
\begin{aligned}
H(M_1, M_2|Z^N) &= H(M_1, M_2|Z^N) - H(M_1, M_2|Y^N) + H(M_1, M_2|Y^N) \\
&= H(M_1, M_2|Z^N) - H(M_1, M_2) + H(M_1, M_2) - H(M_1, M_2|Y^N) + H(M_1, M_2|Y^N) \\
&= I(M_1, M_2; Y^N) - I(M_1, M_2; Z^N) + H(M_1, M_2|Y^N) \\
&\leq I(M_1, M_2; Y^N) - I(M_1, M_2; Z^N) + \delta(P_{e2}) \qquad (80) \\
&\leq \sum_{i=1}^{N} [I(U_i; Y_i) - I(V_i; Z_i|Q_i)] \qquad (81)
\end{aligned}
$$

where Equation (80) is from Fano's inequality.

To serve the single-letter characterization, let us introduce a time-sharing random variable, J, independent of all other random variables and uniformly distributed over $\{1, 2, ..., N\}$. Set:

$$
\begin{aligned}
&U = (U_J, J), V = (V_J, J), Q = (Q_J, J), \\
&A = A_J, B = B_J, S = S_J, X = X_J, Y = Y_J, Z = Z_J
\end{aligned}
$$

Then, substituting the above definition into Equations (78) and (81), the conditions in Equations (17) and (19) are verified straightforwardly. From the definition of the auxiliary random variables, the Markov chains $Q \to V \to U \to Y \to Z$ and $U \to A \to S$ are easy to be verified. We complete the proof of Theorem 4.

Acknowledgments

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

Author Contribution

Xinxing Yin and Zhi Xue did the theoretical work and wrote this paper. All authors have read and approved the final manuscript

Conflicts of Interests

The authors declare no conflict of interest.

References

1. Weissman, T. Capacity of channels with action-dependent states. *IEEE Trans. Inf. Theory* **2010**, *56*, 5396–5411.
2. Asnani, H.; Permuter, H.; Weissman, T. Probing capacity. *IEEE Trans. Inf. Theory* **2011**, *57*, 7317–7332.
3. Steinberg, Y.; Weissman, T. The degraded broadcast channel with action-dependent states. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Boston, MA, USA, 1–6 July 2012; pp. 596–600.
4. Steinberg, Y. The degraded broadcast channel with non-causal action-dependent side information. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, 7–12 July 2013; pp. 2965–2969.
5. Ahmadi, B.; Simeone, O. On channels with action-dependent states. **2012**, arXiv:1202.4438. Available online: http://arxiv.org/abs/1202.4438 (accessed on 22 July 2012).
6. Asnani, H.; Permuter, H.; Weissman, T. To observe or not to observe the channel state. In Proceedings of the Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2010; pp. 1434–1441.
7. Kittichokechai, K.; Oechtering, T.J.; Skoglund, M. Capacity of the channel with action-dependent state and reversible input. In Proceedings of IEEE Swedish Communication Technologies Workshop (Swe-CTW), Stockholm, Swedish, 19–21 October 2011.
8. Kittichokechai, K.; Oechtering, T.J.; Skoglund, M. Coding with action-dependent side information and additional reconstruction requirements. **2012**, arXiv:1202.1484. Available online: http://arxiv.org/abs/1202.1484 (accessed on 7 February 2012).
9. Choudhuri, C.; Mitra, U. Action dependent strictly causal state communication.**2012**, arXiv:1202.0934. Available online: http://arxiv.org/abs/1202.0934 (accessed on 5 February 2012).
10. Choudhuri, C.; Mitra, U. How useful is adaptive action? In Proceedings of the Global Communications Conference, Anaheim, CA, USA, 3–7 December 2012; pp. 2251–2255.

11. Ahmadi, B.; Asnani, H.; Simeone, O.; Permuter, H. Information embedding on actions. In Proceedings of the IEEE IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, 7–12 July 2013; pp. 186–190.
12. Ahmadi, B.; Asnani, H.; Simeone, O.; Permuter, H. Information embedding on actions. **2012**, arXiv:1207.6084. Available online: http://arxiv.org/abs/1207.6084 (accessed on 25 July 2012).
13. Petitcolas, F.A.P.; Anderson, R.J.; Kuhn, M.G. Information hiding—A survey. *Proc. IEEE* **1999**, *87*, 1062–1078.
14. Moulin, P.; O'Sullivan, J.A. Information-theoretic analysis of information hiding. *IEEE Trans. Inf. Theory* **2003**, *49*, 563–593.
15. O'Sullivan, J.A.; Moulin, P.; Ettinger, J.M. Information theoretic analysis of steganography. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA, 16–21 August 1998.
16. Zaidi, A.; Vandendorpe, L. Coding schemes for relay-assisted information embedding. *IEEE Trans. Inf. Forensics Secur.* **2009**, *4*, 70–85.
17. Zaidi, A.; Piantanida, P.; Duhamel, P. Broadcast- and MAC-aware coding strategies for multiple user information embedding. *IEEE Trans. Signal Process.* **2007**, *55*, 2974–2992.
18. Dai, B.; Vinck, A.J.H.; Luo, Y.; Tang, X. Wiretap channel with action-dependent channel state information. *Entropy* **2013**, *15*, 445–473.
19. Dai, B.; Vinck, A.J.H.; Luo, Y. Wiretap channel in the presence of action-dependent states and noiseless feedback. *J. Appl. Math.* **2013**, *2013*, doi:10.1155/2013/423619.
20. Dai, B.; Luo, Y. Some new results on the wiretap channel with side information. *Entropy* **2013**, 14, 1671–1702.
21. Le Treust, M.; Zaidi, A.; Lasaulce S. An achievable rate region for the broadcast wiretap channel with asymmetric side information. In Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 28–30 September 2011; pp. 68–75.
22. Gel'fand, S.I.; Pinsker, M.S. Coding for channel with random parameters. *Probl. Control Inf. Theory* **1980**, *9*, 19–31.
23. Cover, T.M. *Elements of Information Theory*; Wiley: New York, NY, USA,1991.
24. El Gamal, A.; Kim, Y. *Network Information Theory*; Cambridge University Press: New York, NY, USA, 2011.
25. Csiszár, I.; Köner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Academic Press: London, UK, 1981.