Wiretap Channel With Causal State Information

Yeow-Khiang Chia and Abbas El Gamal, *Fellow, IEEE*

***Abstract*—A lower bound on the secrecy capacity of the wiretap channel with state information available causally at both the encoder and the decoder is established. The lower bound is shown to be strictly larger than that for the noncausal case by Liu and Chen. Achievability is proved using block Markov coding, Shannon strategy, and key generation from common state information. The state sequence available at the end of each block is used to generate a key to enhance the transmission rate of the confidential message in the following block. An upper bound on the secrecy capacity when the state is available noncausally at the encoder and the decoder is established and is shown to coincide with the aforementioned lower bound for several classes of wiretap channels with state.**

***Index Terms*—Channels with state, secrecy capacity, wiretap channel.**

I. Introduction

CONSIDER the two-receiver wiretap channel with state depicted in Fig. 1. The sender X wishes to communicate a message reliably to the legitimate receiver Y while keeping it asymptotically secret from the eavesdropper Z. The secrecy capacity for this channel can be defined under various scenarios of state information availability at the encoder and the decoder. When the state information is not available at either party, the problem reduces to the classical wiretap channel for the channel averaged over the state and the secrecy capacity is known [1], [2]. When the state is available only at the decoder, the problem reduces to the wiretap channel with augmented receiver (Y, S). The interesting cases to consider, therefore, are when the state information is available at the encoder and may or may not be available at the decoder. This raises the question of how the encoder and the decoder can make use of the state information to increase the secrecy rate. This model is a generalization of the wiretap channel with shared secret key in [3] and can be used also as a base model for secret communication over fast-fading channels in which the sender and the receiver have some means for measuring the channel statistics but the eavesdropper does not. In [4], Chen and Vinck established a lower bound on the secrecy capacity when the state information is available noncausally only at the encoder. The lower bound is established using a combination of Gelfand–Pinsker coding and Wyner wiretap coding. Subsequently, Liu and Chen [5] used the same techniques to establish a lower bound on the secrecy capacity when the state information is available noncausally at both the encoder and the decoder. In a related direction, Khisti *et al.* [6] considered the problem of secret key agreement first studied in [7] and [8] for the wiretap channel with state and established the secret key capacity when the state is available causally or noncausally at the encoder and the decoder. The key is generated in two parts: the first using a wiretap channel code while treating the state sequence as a time-sharing sequence, and the second is generated from the state itself.

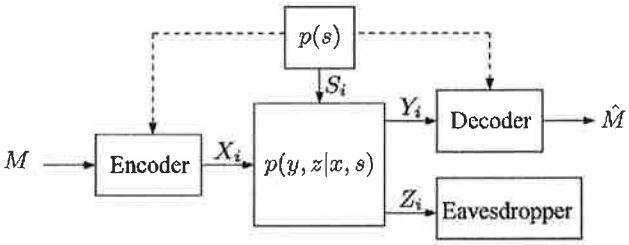

Fig. 1. Wiretap channel with state.

In this paper, we consider the wiretap channel with state information available *causally* at the encoder and the decoder. We establish a lower bound on the secrecy capacity, which is strictly larger than the lower bound for the noncausal case in [5]. Our achievability scheme, however, is quite different from the scheme in [5]. We use block Markov coding, Shannon strategy for channels with state [9], and secret key agreement from state information, which builds on the work in [6]. However, unlike [6], we are not directly interested in the size of the secret key, but rather in using the secret key generated from the state sequence in one transmission block to increase the secrecy rate in the following block. This block Markov scheme causes additional information leakage through the correlation between the secret key generated in a block and the received sequences at the eavesdropper in subsequent blocks. We show that this leakage is asymptotically negligible. Although a similar block Markov coding scheme was used in [10] to establish the secrecy capacity of the degraded wiretap channel with rate limited secure feedback, in their setup no information about the key is leaked to the eavesdropper because the feedback link is assumed to be secure. We also establish an upper bound on the secrecy capacity of the wiretap channel with state information available noncausally at the encoder and decoder. We show that the upper bound coincides with the aforementioned lower bound for several classes of channels. Thus, the secrecy capacity for these classes do not depend on whether the state information is known causally or noncausally at the encoder.

The rest of this paper is organized as follows. In Section II, we provide the needed definitions. In Section III, we summarize and discuss the main results in this paper. The proofs of the lower and upper bounds are detailed in Sections IV and V, respectively.

Manuscript received August 31, 2010; revised June 18, 2011; accepted August 01, 2011. Date of publication January 23, 2012; date of current version April 17, 2012. The material in this paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: ykchia@stanford.edu; abbas@ee.stanford.edu).

Communicated by M. Gastpar, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2011.2181329

II. Problem Definition

Consider a discrete memoryless wiretap channel (DM-WTC) with discrete memoryless state $(\mathcal{X} \times \mathcal{S}, p(y,z|x,s)p(s), \mathcal{Y}, \mathcal{Z})$ that consists of a finite input alphabet $\mathcal{X}$, finite output alphabets $\mathcal{Y}$, $\mathcal{Z}$, a finite state alphabet $\mathcal{S}$, a collection of conditional probability mass functions (pmfs) $p(y,z|x,s)$ on $\mathcal{Y} \times \mathcal{Z}$, and a pmf $p(s)$ on the state alphabet $\mathcal{S}$. The sender X wishes to send a confidential message $M \in [1 : 2^{nR}]$ to the receiver Y while keeping it secret from the eavesdropper Z with either causal or noncausal state information available at both the encoder and decoder.

A $(2^{nR}, n)$ code for the DM-WTC with causal state information at the encoder and decoder consists of the following: a message set $[1 : 2^{nR}]$; an encoder that generates a symbol $X_i(m)$ according to a conditional pmf $p(x_i|m, s^i, x^{i-1})$ for $i \in [1 : n]$; and a decoder that assigns an estimate $\hat{M}$ or an error message to each received sequence pair (y^n, s^n). We assume that the message M is uniformly distributed over the message set. The probability of error is defined as $P_e^{(n)} = P\{\hat{M} \neq M\}$. The *information leakage rate* at the eavesdropper Z, which measures the amount of information about M that leaks out to the eavesdropper, is defined as $R_L = \frac{1}{n} I(M; Z^n)$. A secrecy rate R is said to be achievable if there exists a sequence of codes with $P_e^{(n)} \to 0$ and $R_L \to 0$ as $n \to \infty$. The *secrecy capacity* $C_{\text{S-CSI}}$ is the supremum of the set of achievable rates.

We also consider the case when the state information is available noncausally at the encoder. The only change in the aforementioned definitions is that the encoder now generates a codeword $X^n(m)$ according to a conditional pmf $p(x^n|m, s^n)$, i.e., a random mapping that depends on the entire state sequence instead of just the past and present state sequence. The secrecy capacity for this scenario is denoted by $C_{\text{S-NCSI}}$.

The notation used in this paper will follow that in [11].

III. Summary of Main Results

We present the results in this paper. The proofs of these results are given in the following two sections and in the Appendix.

A. Lower Bound

The main result in this paper is the following lower bound on the secrecy capacity of the DM-WTC with state information available causally at both the encoder and decoder.

Theorem 1: The secrecy capacity of the DM-WTC with state information available causally at the encoder and decoder is lower bounded as

$$\begin{aligned} C_{\text{S-CSI}} &\geq \max \Bigg\{ \max_{P_1} \min \{I(U;Y,S) - I(U;Z,S) \\ &\qquad + H(S|Z), I(U;Y,S)\}, \\ &\qquad \max_{P_2} \min\{H(S|Z,V), I(V;Y|S)\} \Bigg\} \end{aligned} \tag{1}$$

where P_1 is of the form $p(u), v(u,s), p(x|v,s)$ and P_2 is of the form $p(v)p(x|v,s)$.

Note that if $S = \emptyset$, the aforementioned lower bound reduces to the secrecy capacity for the wiretap channel. Clearly, this lower bound also holds when noncausal state information is available at the encoder, since we can always treat the noncausal state information as causal state information. Define

$$\begin{aligned} R_{\text{S-CSI-1}} &= \max_{p(u),v(u,s),p(x|v,s)} \min \{I(U;Y,S) - I(U;Z,S) \\ &\qquad + H(S|Z), I(U;Y,S)\} \\ R_{\text{S-CSI-2}} &= \max_{p(v)p(x|v,s)} \min\{H(S|Z,V), I(V;Y|S)\}. \end{aligned} \tag{2}$$

Then, (1) can be expressed as

$$C_{\text{S-CSI}} \geq \max\{R_{\text{S-CSI-1}}, R_{\text{S-CSI-2}}.\}$$

The proof of this theorem is detailed in Section IV.

Remark 3.1: Using the functional representation lemma [12], $R_{\text{S-CSI-1}}$ can be equivalently written as

$$\begin{aligned} R_{\text{S-CSI-1}} &= \max_{p(v|s)p(x|v,s)} \min \{I(V;Y|S) - I(V;Z|S) \\ &\qquad + H(S|Z), I(V;Y|S)\}. \end{aligned} \tag{3}$$

Unless otherwise stated, this equivalent characterization for $R_{\text{S-CSI-1}}$ will be assumed for the rest of this section to derive other results. From Section IV onward, we revert back to the original characterization in (2).

In [5], the authors established the following lower bound for the noncausal case:

$$\begin{aligned} C_{\text{S-NCSI}} &\geq \max_{p(u|s)p(x|u,s)} (I(U;Y,S) - \max\{I(U;Z), I(U;S)\}) \\ &= \max_{p(u|s)p(x|u,s)} \min \{I(U;Y|S) - I(U;Z|S) \\ &\qquad + I(S;U|Z), I(U;Y|S)\}. \end{aligned} \tag{4}$$

From (3), $R_{\text{S-CSI-1}}$ is clearly at least as large as this lower bound. Hence, our lower bound (1) is at least as large as this lower bound (4). We now show that the lower bound (4) is as large as $R_{\text{S-CSI-1}}$.

Fix $V \in [0 : |\mathcal{V}| - 1]$, $p(v|s)$, and $p(x|v,s)$ in $R_{\text{S-CSI-1}}$. Let $U \in [0 : |\mathcal{V}||\mathcal{S}| - 1]$ in bound (4). Define the conditional pmfs: for $u = v + s|\mathcal{V}|$, let $p(u|s) = p(v|s)$, $p(x|u,s) = p(x|v,s)$, and let $p(u|s) = p(x|u,s) = 0$ otherwise. Under this mapping, it is easy to see that $H(S|Z,U) = 0$ and the other terms in (4) reduce to those in $R_{\text{S-CSI-1}}$.

The following shows that our lower bound (1) can be strictly larger than (4).

Example: Consider the channel in Fig. 2, where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S} \in \{0,1\}$, $p(y,z|x,s) = p(y,z|x)$, and the conditional pmf defined in the figure. The state S with entropy $H(S) = 1 - H(0.1)$ is observed by both X and Y.

By setting $V = X$ independent of S and $P\{X = 1\} = P\{X = 0\} = 0.5$ in $R_{\text{S-CSI-2}}$, we obtain $R_{\text{S-CSI-2}} \geq 1 - H(0.1)$.

We now show that $R_{\text{S-CSI-1}}$ is strictly smaller than $1 - H(0.1)$. First, note that

$$\begin{aligned} I(V;Y|S) &= H(Y|S) - H(Y|V,S) \\ &\le H(Y) - H(Y|X) \\ &= I(X;Y) \le 1 - H(0.1). \end{aligned}$$

However, for $R_{\text{S-CSI-1}} \ge 1 - H(0.1)$, we must have $I(V;Y|S) \ge 1 - H(0.1)$. Hence, we must have $I(V;Y|S) = 1 - H(0.1)$. Next, consider

$$\begin{aligned} I(V;Y|S) &= H(Y|S) - H(Y|V,S) \\ &\overset{(a)}{\le} 1 - H(Y|V,S) \\ &\overset{(b)}{\le} 1 - H(Y|V,S,X) \\ &= 1 - H(0.1). \end{aligned}$$

Step (a) holds with equality iff $p(y|s) = 0.5$ for all $y, s \in \{0,1\}$. From the structure of the channel, this implies that $p(x|s) = 0.5$ for all $x, s \in \{0,1\}$. Step (b) holds with equality iff $H(Y|X,V,S) = H(Y|V,S)$, or equivalently $I(X;Y|V,S) = 0$. This implies that given V, S, X and Y are independent, $p(x,y|v,s) = p(x|v,s)p(y|v,s)$. But since $p(x,y|v,s) = p(x|v,s)p(y|x)$, either 1) $p(x|v,s) = 0$ or 2) $p(y|v,s) = p(y|x)$ must hold. Now, consider the pair $x = 1, y = 1$. Then, we must have either 1) $p(x = 1|v,s) = 0$ or 2) $p(y = 1|v,s) = p(y = 1|x = 1) = 0.9$. In 1), X is a function of V and S. In 2), we have

$$\begin{aligned} p(y = 1|v,s) &= p(x = 1|v,s)p(y = 1|x = 1) \\ &\quad + (1 - p(x = 1|v,s))p(y = 1|x = 0) \\ &= 0.9p(x = 1|v,s) + 0.1 \\ &\quad - 0.1p(x = 1|v,s) \\ &= 0.8p(x = 1|v,s) + 0.1. \end{aligned}$$

Since $p(y = 1|v,s) = 0.9$, we have $0.8p(x = 1|v,s) + 0.1 = 0.9$, which implies that $p(x = 1|v,s) = 1$. Here again X is a function of V and S. Hence, in both cases 1) and 2), X must be a function of V and S, which implies that $Z = X$ is also a function of V and S. Using the fact that $p(x|s) = p(z|s) = 0.5$ for all x, s, we have

$$\begin{aligned} I(V;Z|S) &= H(Z|S) - H(Z|V,S) \\ &= H(X|S) \\ &= 1. \end{aligned}$$

The first expression in $R_{\text{S-CSI-1}}$ is, then, upper bounded by

$$\begin{aligned} &I(V;Y|S) - I(V;Z|S) + H(S|Z) \\ &\le I(V;Y|S) - I(V;Z|S) + H(S) \\ &= 1 - H(0.1) - 1 + 1 - H(0.1) \\ &= 1 - 2H(0.1) < 1 - H(0.1). \end{aligned}$$

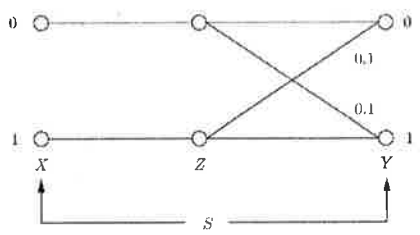

Fig. 2. Example.

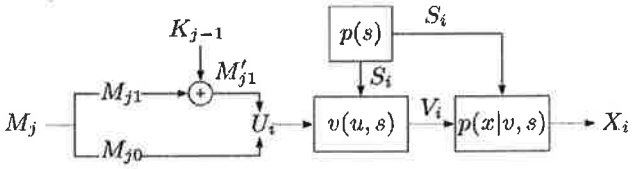

Fig. 3. Encoding in block j.

This shows that $R_{\text{S-CSI-1}} < R_{\text{S-CSI-2}}$.

The proof of Theorem 1 is detailed in Section IV. To illustrate the main ideas, we outline the proof for the characterization of $R_{S-CSI-1}$ in (2) for the case when $I(U;Y,S) - I(U;Z,S) > 0$. Our coding scheme involves the transmission of $b - 1$ independent messages over b n-transmission blocks. We split each message M_j, $j \in [2 : b]$, into two independent messages $M_{j0} \in [1 : 2^{nR_0}]$ and $M_{j1} \in [1 : 2^{nR_1}]$ with $R_0 + R_1 = R$. Codebook generation consists of two steps. The first is the generation of the *message codebook*. We randomly generate sequences $u^n(l)$, $l \in [1 : 2^{nI(U;Y,S)}]$, and partition the set of indices $[1 : 2^{nI(U;Y,S)}]$ into 2^{nR_0} equal size bins. The indices in each bin are further partitioned into 2^{nR_1} equal size sub-bins $\mathcal{C}(m_0, m_1)$. The second step is to generate the key codebook. We randomly bin the set of state sequences s^n into 2^{nR_K} bins $\mathcal{B}(k)$. The key K_{j-1} used in block j is the bin index of the state sequence $\mathbf{S}(j-1)$ in block $j-1$.

To send message M_j, M_{j1} is encrypted using the key K_{j-1} to obtain $M'_{j1} = M_{j1} \oplus K_{j-1}$. A codeword $u^n(L)$ is selected uniformly at random from sub-bin $\mathcal{C}(M_{j0}, M_{j1} \oplus K_{j-1})$ and transmitted using Shannon's strategy as depicted in Fig. 3. The decoder uses joint typicality decoding together with its knowledge of the key to decode message M_j at the end of block j. Finally, at the end of block j, the encoder and the decoder declare the bin index K_j of the state sequence $\mathbf{s}(j)$ as the key to be used in block $j+1$. To show that the messages can be kept asymptotically secret from the eavesdropper, note that M_{j0} is transmitted using Wyner wiretap coding. Hence, it can be kept secret from the eavesdropper provided $I(U;Y,S) - I(U;Z,S) > 0$. The key part of the proof is to show that the second part of the message M_{j1}, which is encrypted with the key K_{j-1}, can be kept secret from the eavesdropper. This involves showing that the eavesdropper has negligible information about K_{j-1}. However, the fact that K_{j-1} is generated from the state sequence in block $j-1$ and used in block j results in correlation between it and all received sequences at the eavesdropper from subsequent blocks. We show that if $R_K < H(S|Z)$, then the eavesdropper has negligible information about K_{j-1} given all its received sequences.

B. Upper Bound

We establish the following upper bound on the secrecy capacity of the wiretap channel with noncausal state information available at both the encoder and decoder (which holds also for the causal case).

Theorem 2: The following is an upper bound to the secrecy capacity of the DM-WTC with state noncausally available at the encoder and decoder:

$$C_{\text{S-NCSI}} \le \min\{I(V_1;Y|U,S) - I(V_1;Z|U,S) + H(S|Z,U), I(V_2;Y|S)\}$$

for some U, V_1, and V_2 such that $p(u,v_1,v_2,x|s) = p(u|s)p(v_1|u,s)p(v_2|v_1,s)p(x|v_2,s)$. The cardinality of the auxiliary random variables can be upper bounded by $|\mathcal{U}| \le |\mathcal{S}|(|\mathcal{X}|+1)$, $|\mathcal{V}_1| \le |\mathcal{U}||\mathcal{S}|(|\mathcal{X}|+1)$ and $|\mathcal{V}_2| \le |\mathcal{V}_1||\mathcal{U}||\mathcal{S}||\mathcal{X}|$.

The proof of this theorem is given in Section V.

C. Secrecy Capacity Results

We show that the upper bound in Theorem 2 coincides with the lower bound in Theorem 1 for the following cases.

1) Class of Less Noisy Channels: We show that Theorems 1 and 2 are also tight when $I(U;Y|S) \ge I(U;Z|S)$ for every U such that $(U,S) \to (X,S) \to (Y,Z)$ form a Markov chain, i.e., when Y is *less noisy* than Z (see [13]) for every state $s \in \mathcal{S}$ [13].

Theorem 3: The secrecy capacity for the DM-WTC with the state information available causally or noncausally at the encoder and decoder when Y is less noisy than Z is

$$\begin{aligned} C_{\text{S-CSI}} &= C_{\text{S-NCSI}} \\ &= \max_{p(x|s)} \min\{I(X;Y|S) - I(X;Z|S) + H(S|Z), I(X;Y|S)\}. \end{aligned}$$

Consider the special case when $p(y,z|x,s) = p(y,z|x)$ and Z is a *degraded* version of Y; then, Theorem 3 specializes to the secrecy capacity for the wiretap channel with a key [3]

$$\begin{aligned} C_{\text{S-CSI}} &= C_{\text{S-NCSI}} \\ &= \max_{p(x)} \min\{I(X;Y) - I(X;Z) + H(S), I(X;Y)\}. \end{aligned}$$

Achievability for Theorem 3 follows directly from Theorem 1 by setting $V = X$ and observing $R_{\text{S-CSI-1}} \ge R_{\text{S-CSI-2}}$ since Y is less noisy than Z. Hence, the achievability scheme for $R_{\text{S-CSI-1}}$ is optimal for this class of channels. To establish the converse, we use the less noisy assumption to strengthen the first inequality in Theorem 2 as follows:

$$\begin{aligned} &I(V_1;Y|U,S) - I(V_1;Z|U,S) + H(S|Z,U) \\ &\le I(V_1;Y|U,S) - I(V_1;Z|U,S) + H(S|Z) \\ &\overset{(a)}{\le} I(V_1;Y|S) - I(V_1;Z|S) + H(S|Z) \\ &\overset{(b)}{\le} I(X;Y|S) - I(X;Z|S) + H(S|Z) \end{aligned}$$

where $(a)$and(b) follow from the less noisy assumption. The proof of the second inequality follows by the data processing inequality, which yields $I(V_2;Y|S) \le I(X;Y|S)$.

2) Channel is Independent of State and Eavesdropper Is Less Noisy Than Receiver: Next, consider the case where $p(y,z|x,s) = p(y,z|x)$ and the eavesdropper Z is less noisy than Y, that is, $I(U;Z) \ge I(U;Y)$ for every U such that $U \to X \to (Y,Z)$. Then, the capacity of this special class of channels is

$$C_{\text{S-CSI}} = C_{\text{S-NCSI}} = \max_{p(x)} \min\{H(S), I(X;Y)\}.$$

Achievability follows by setting $V = X$ independent of S. We note that the scheme here is basically a "one-time pad" scheme where the message that is transmitted is scrambled with a key generated from the state sequence. Here, the secrecy capacity is achieved by $R_{\text{S-CSI-2}}$, and hence, $R_{\text{S-CSI-2}} \ge R_{\text{S-CSI-1}}$. The example we gave to illustrate the fact that $R_{\text{S-CSI-2}}$ can be larger than $R_{\text{S-CSI-1}}$ is a special case of this class of channels. The converse follows from Theorem 2 and the observation that since Z is less noisy than Y and $p(y,z|x,s) = p(y,z|x)$

$$\begin{aligned} &I(V_1;Y|U,S) - I(V_1;Z|U,S) + H(S|Z,U) \\ &\quad \le H(S|Z,U) \\ &\quad \le H(S) \end{aligned}$$

and $I(V_2;Y|S) \le I(X;Y)$.

3) Specific Mutual Information Constraints: Following the lines of [4], we can show that Theorems 1 and 2 are tight for the following two special cases. The achievability scheme for $R_{S-CSI-1}$ is optimal for both special cases.

1) If there exists a V^* such that $\max_{p(v|s)p(x|v,s)}(I(V;Y|S) - I(V;Z|S) + H(S|Z)) = I(V^*;Y|S) - I(V^*;Z|S) + H(S|Z)$ and $I(V^*;Y|S) - I(V^*;Z|S) + H(S|Z) \le I(V^*;Y|S)$, then the secrecy capacity is $C_{\text{S-CSI}} = C_{\text{S-NCSI}} = I(V^*;Y|S) - I(V^*;Z|S) + H(S|Z)$.
2) If there exists a V' such that $\max_{p(v|s)p(x|v,s)} I(V;Y|S) = I(V';Y|S)$ and $I(V';Y|S) \le I(V';Y|S) - I(V';Z|S) + H(S|Z)$, then the secrecy capacity is $C_{\text{S-CSI}} = C_{\text{S-NCSI}} = I(V';Y|S)$.

IV. Proof of Theorem 1

We prove achievability of $R_{\text{S-CSI-1}}$ and $R_{\text{S-CSI-2}}$ separately. The proof for $R_{\text{S-CSI-1}}$ is split into Cases 1 and 2 while $R_{\text{S-CSI-2}}$ is proved in Case 3.

Case 1: $R_{\text{S-CSI-1}}$ With $I(U;Y,S) > I(U;Z,S)$:

Codebook Generation: Split message M_j into two independent messages $M_{j0} \in [1 : 2^{nR_0}]$ and $M_{j1} \in [1 : 2^{nR_1}]$; thus, $R = R_0 + R_1$. Let $\tilde{R} \ge R$. The codebook generation consists of two parts.

Message Codeword Generation: Randomly and independently generate sequences $u^n(l)$, $l \in [1 : 2^{n\tilde{R}}]$, each according to $\prod_{i=1}^n p_U(u_i)$ and partition the set of indices $[1 : 2^{n\tilde{R}}]$ into 2^{nR_0} equal-size bins $\mathcal{C}(m_0)$, $m_0 \in [1 : 2^{nR_0}]$. Further partition the indices within each bin $\mathcal{C}(m_0)$ into 2^{nR_1} equal size

sub-bins, $\mathcal{C}(m_0, m_1)$, $m_1 \in [1 : 2^{nR_1}]$. Hence, $l \in \mathcal{C}(m_0, m_1)$ if and only if $(m_0 + m_1 - 1 - 1)2^{n(\tilde{R}-R_0-R_1)} + 1 \leq l \leq (m_0 + m_1)2^{n(\tilde{R}-R_0-R_1)}$.

Key Codebook Generation: Randomly and uniformly partition the set of s^n sequences into 2^{nR_K} bins $\mathcal{B}(k)$, $k \in [1 : 2^{nR_K}]$.

Both codebooks are revealed to all parties.

Encoding: We send $b-1$ messages over b n-transmission blocks. In the first block, we randomly select an index $L \in \mathcal{C}(m_{10}, m'_{11})$. The encoder, then, computes $v_i = v(u_i(L), s_i)$ and transmits a randomly generated symbol $X_i \sim p(x_i|s_i, v_i)$ for $i \in [1 : n]$. At the end of the first block, the encoder and the decoder declare $k_1 \in [1 : 2^{nR_K}]$ such that $\mathbf{s}(1) \in \mathcal{B}(k_1)$ as the key to be used in block 2.

Encoding in block $j \in [2 : b]$ proceeds as follows. To send message $m_j = (m_{j0}, m_{j1})$ and given key k_{j-1}, the encoder computes $m'_{j1} = m_{j1} \oplus k_{j-1}$. To ensure secrecy, we must have $R_1 \leq R_K$ [14]. The encoder then randomly selects an index L such that $L \in \mathcal{C}(m_{j0}, m'_{j1})$. It then computes $v_i = v(u_i(L), s_i)$ and transmits a randomly generated symbol $X_i \sim p(x_i|s_i, v_i)$ for $i \in [(j-1)n + 1 : jn]$.

Decoding and Analysis of the Probability of Error: At the end of block j, the decoder declares that $\hat{l}$ is sent if it is the unique index such that $(u^n(\hat{l}), \mathbf{Y}(j), \mathbf{S}(j)) \in \mathcal{T}_\epsilon^{(n)}$, otherwise it declares an error. It then finds the index pair $(\hat{m}_{j0}, \hat{m}'_{j1})$ such that $\hat{l} \in \mathcal{C}(\hat{m}_{j0}, \hat{m}'_{j1})$. Finally, it recovers $\hat{m}_{j1}$ by computing $\hat{m}_{j1} = (\hat{m}'_{j1} - k_{j-1})_{\mathrm{mod}\, 2^{nR_K}}$.

To analyze the probability of error, let $\epsilon'' > \epsilon' > \epsilon > 0$ and define the following events for every $j \in [2 : b]$:

$$\begin{aligned}
\mathcal{E}(j) &= \{\hat{M}_j \neq M_j\} \\
\mathcal{E}_1(j) &= \{(U^n(L), \mathbf{S}(j)) \notin \mathcal{T}_{\epsilon'}^n\} \\
\mathcal{E}_2(j) &= \{(U^n(L), \mathbf{S}(j), \mathbf{Y}(j)) \notin \mathcal{T}_{\epsilon''}^n\} \\
\mathcal{E}_3(j) &= \{(U^n(\hat{l}), \mathbf{S}(j), \mathbf{Y}(j)) \in \mathcal{T}_{\epsilon''}^n \text{ for some } \hat{l} \neq L\}.
\end{aligned}$$

The probability of error is upper bounded as

$$P(\mathcal{E}) = P\{\cup_{j=2}^b \mathcal{E}(j)\} \leq \sum_{j=2}^b P(\mathcal{E}(j)).$$

Each probability of error term can be upper bounded as

$$\begin{aligned}
P(\mathcal{E}(j)) \leq P(\mathcal{E}_1(j)) + P(\mathcal{E}_2(j) \cap \mathcal{E}_1^c(j)) \\
+ P(\mathcal{E}_3(j) \cap \mathcal{E}_2^c(j)).
\end{aligned}$$

Now, $P(\mathcal{E}_1(j)) \to 0$ as $n \to \infty$ by the law of large numbers (LLN) since $P\{(U^n(L) \in \mathcal{T}_\epsilon^{(n)})\} \to 1$ as $n \to \infty$ and $\mathbf{S}(j) \sim \prod_{i=1}^n p_S(s_i) = \prod_{i=1}^n p_{S|U}(s_i|u_i)$ by independence. The term $P(\mathcal{E}_2(j) \cap \mathcal{E}_1^c(j)) \to 0$ as $n \to \infty$ by LLN since $(U^n(L), \mathbf{S}(j) \in \mathcal{T}_{\epsilon'}^n$ and $Y^n \sim \prod_{i=1}^n pY|U, S(y_i|u_i, s_i)$. For the last term, consider

$$P(\mathcal{E}_3(j) \cap \mathcal{E}_2^c(j)) = \sum_l p(l) P(\mathcal{E}_3(j) \cap \mathcal{E}_2^c(j)|L = l).$$

Note that L is independent of the transmission codebook sequences $(u^n(l), l \in [1 : 2^{n\tilde{R}}])$ and the current state sequence $\mathbf{S}(j)$. Therefore, by the packing lemma [11, Lecture 3], $P(\mathcal{E}_3(j) \cap \mathcal{E}_2^c(j)|L = l) \to 0$ as $n \to \infty$ if $\tilde{R} < I(U; Y, S) - \delta(\epsilon'')$. Hence, $P(\mathcal{E}_3(j) \cap \mathcal{E}_2^c(j)) \to 0$ as $n \to \infty$ if $\tilde{R} < I(U; Y, S) - \delta(\epsilon'')$.

Analysis of the Information Leakage Rate: Let $\mathbf{Z}(j)$ denote the eavesdropper's received sequence in block $j \in [1 : b]$ and $\mathbf{Z}^j = (\mathbf{Z}(1), \ldots, \mathbf{Z}(j))$. We will need the following two results.

Proposition 1: If $R_K < H(S|Z) - 4\delta(\epsilon)$ and $\tilde{R} \geq I(U; Z, S) + \delta(\epsilon)$, then the following holds for every $j \in [1 : b]$.

1) $H(K_j|\mathcal{C}) \geq n(R_K - \delta(\epsilon))$.
2) $I(K_j; \mathbf{Z}(j)|\mathcal{C}) \leq n(\delta'(\epsilon) + \delta''(\epsilon))$.
3) $I(K_j; \mathbf{Z}^j|\mathcal{C}) \leq n\delta'''(\epsilon)$, where $\delta'(\epsilon) \to 0$, $\delta''(\epsilon) \to 0$, and $\delta'''(\epsilon) \to 0$ as $\epsilon \to 0$.

The proof of this proposition is given in Appendix A.

Lemma 1: Let $(U, V, Z) \sim p(u, v, z)$, $\bar{R} \geq 0$, and $\epsilon > 0$. Let U^n be a random sequence distributed according to $\prod_{i=1}^n p_U(u_i)$. Let $V^n(l)$, $l \in [1 : 2^{n\bar{R}}]$, be a set of random sequences that are conditionally independent given U^n and each distributed according to $\prod_{i=1}^n p_{V|U}(v_i|u_i)$. Define $\mathcal{C} = \{U^n, V^n(l)\}$. Let $L \in [1 : 2^{n\bar{R}}]$ be a random index distributed according to an arbitrary pmf. Then, if $P\{(U^n, V^n(L), Z^n) \in \mathcal{T}_\epsilon^{(n)}\} \to 1$ as $n \to \infty$ and $\bar{R} > I(V; Z|U) + \delta(\epsilon)$, there exists a $\delta'(\epsilon) > 0$, where $\delta'(\epsilon) \to 0$ as $\epsilon \to 0$, such that, for n sufficiently large, $H(L|Z^n, U^n, \mathcal{C}) \leq n(\bar{R} - I(V; Z|U)) + n\delta'(\epsilon)$.

This lemma is proved in [15].

We are now ready to upper bound the leakage rate averaged over codes. Consider

$$\begin{aligned}
&I(M_2, M_3, \ldots, M_b; \mathbf{Z}^b|\mathcal{C}) \\
&= \sum_{j=2}^b I(M_j; \mathbf{Z}^b|\mathcal{C}, M_{j+1}^b) \\
&\overset{(a)}{\leq} \sum_{j=2}^b I(M_j; \mathbf{Z}^b|\mathcal{C}, \mathbf{S}(j), M_{j+1}^b) \\
&\overset{(b)}{=} \sum_{j=2}^b I(M_j; \mathbf{Z}^j|\mathcal{C}, \mathbf{S}(j))
\end{aligned}$$

where (a) follows by the independence of M_j and $(\mathbf{S}(j), M_{j+1}^b)$, and (b) follows by the Markov Chain relationship $(\mathbf{Z}_{j+1}^b, M_{j+1}^b, \mathcal{C}) \to (\mathbf{Z}^j, \mathbf{S}(j), \mathcal{C}) \to (M_j, \mathcal{C})$. Hence, it suffices to upper bound each individual term $I(M_j; \mathbf{Z}^j|\mathcal{C}, \mathbf{S}(j))$. Consider

$$\begin{aligned}
&I(M_j; \mathbf{Z}^j|\mathcal{C}, \mathbf{S}(j)) \\
&= I(M_{j0}, M_{j1}; \mathbf{Z}^j|\mathcal{C}, \mathbf{S}(j)) \\
&= I(M_{j0}, M_{j1}; \mathbf{Z}^{j-1}|\mathcal{C}, \mathbf{S}(j)) \\
&\quad + I(M_{j0}, M_{j1}; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}).
\end{aligned}$$

Note that the first term is equal to zero by the independence of M_j and past transmissions, the codebook, and state sequence. For the second term, we have

$$\begin{aligned}
&I(M_{j0}, M_{j1}; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}) \\
&= I(M_{j0}; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}) \\
&\quad + I(M_{j1}; \mathbf{Z}(j)|\mathcal{C}, M_{j0}, \mathbf{S}(j), \mathbf{Z}^{j-1}).
\end{aligned}$$

We now bound each term separately. Consider the first term

$$\begin{aligned}
&I(M_{j0};\mathbf{Z}(j)|\mathcal{C},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&= I(M_{j0},L;\mathbf{Z}(j)|\mathcal{C},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad - I(L;\mathbf{Z}(j)|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\le I(U^n;\mathbf{Z}(j)|\mathcal{C},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad - H(L|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + H(L|\mathcal{C},\mathbf{Z}(j),M_{j0},\mathbf{S}(j))\\
&\le \sum_{i=1}^{n}(H(\mathbf{Z}_i(j)|\mathcal{C},\mathbf{S}_i(j)) - H(\mathbf{Z}_i(j)|\mathcal{C},U_i,\mathbf{S}_i(j)))\\
&\quad - H(L|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + H(L|\mathcal{C},\mathbf{Z}(j),M_{j0},\mathbf{S}(j))\\
&\overset{(a)}{\le} nI(U;Z|S) - H(L|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + H(L|\mathcal{C},\mathbf{Z}(j),M_{j0},\mathbf{S}(j))\\
&\overset{(b)}{\le} nI(U;Z|S) - H(L|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + n(\tilde{R} - R_0 - I(U;Z,S) + \delta'(\epsilon))\\
&\overset{(c)}{=} n(\tilde{R} - R_0) - H(L|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1}) + n\delta'(\epsilon)\\
&= n(\tilde{R} - R_0) - H(M_{j1}\oplus K_{j-1}|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad - H(L|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1},M_{j1}\oplus K_{j-1}) + n\delta'(\epsilon)\\
&\le n(\tilde{R} - R_0)\\
&\quad - H(M_{j1}\oplus K_{j-1}|\mathcal{C}.M_{j0},\mathbf{S}(j),K_{j-1},\mathbf{Z}^{j-1})\\
&\quad - n(\tilde{R} - R_0 - R_K) + n\delta'(\epsilon)\\
&\overset{(d)}{=} nR_K - H(M_{j1}\oplus K_{j-1}|\mathcal{C},M_{j0},\mathbf{S}(j),K_{j-1})\\
&\quad + n\delta'(\epsilon)\\
&= nR_K - H(M_{j1}|\mathcal{C},M_{j0},\mathbf{S}(j),K_{j-1}) + n\delta'(\epsilon)\\
&= n\delta'(\epsilon)
\end{aligned}$$

where (a) follows from the fact that $H(\mathbf{Z}_i(j)|\mathcal{C},\mathbf{S}_i(j)) \le H(\mathbf{Z}_i(j)|\mathbf{S}_i(j)) = H(Z|S)$ and $H(\mathbf{Z}_i(j)|\mathcal{C},U_i,\mathbf{S}_i(j)) = H(Z|U,S)$. Step (b) follows by Lemma 1 which requires that 1) $P\{(U^n(L),\mathbf{S}(j),\mathbf{Z}(j)) \in \mathcal{T}_\epsilon^{(n)}\} \to 1$ as $n\to\infty$, and 2) $\tilde{R} - R_0 > I(U;Z,S) + \delta(\epsilon)$; 1) can be established using the same steps as in the analysis of probability of error. Step (c) follows by the independence of U and S. Step (d) follows by the Markov Chain relationship

$$\begin{aligned}
(\mathbf{Z}^{j-1},M_{j0},\mathbf{S}(j)) &\to (K_{j-1},M_{j0},\mathbf{S}(j))\\
&\to (M_{j1}\oplus K_{j-1},M_{j0},\mathbf{S}(j)).
\end{aligned}$$

The last step follows by the fact that M_{j1} is independent of $(\mathcal{C},M_{j0},\mathbf{S}(j),K_{j-1})$ and uniformly distributed over $[1:2^{nR_K}]$.

Next, consider the second term

$$\begin{aligned}
&I(M_{j1};\mathbf{Z}(j)|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\le I(M_{j1},L;\mathbf{Z}(j)|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad - I(L;\mathbf{Z}(j)|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\le I(U^n;\mathbf{Z}(j)|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad - H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j})\\
&\overset{(a)}{\le} nI(U;Z|S) - H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j})\\
&\overset{(b)}{\le} nI(U;Z|S) - H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + n(\tilde{R} - R_0) - nI(U;Z,S) + n\delta'(\epsilon)\\
&= n(\tilde{R} - R_0)\\
&\quad - H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1}) + n\delta'(\epsilon)
\end{aligned}$$

where (a) follows from the same steps used in bounding $I(M_{j0};\mathbf{Z}(j)|\mathcal{C},\mathbf{S}(j),\mathbf{Z}^{j-1})$ and (b) follows from Lemma 1, which requires the same condition $\tilde{R} - R_0 > I(U;Z,S) + \delta(\epsilon)$. Next consider

$$\begin{aligned}
&H(L|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&= H(M_{j1}\oplus K_{j-1}|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + H(L|\mathcal{C},M_{j0},M_{j1},M_{j1}\oplus K_{j-1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&= H(K_{j-1}|\mathcal{C},M_{j0},M_{j1},\mathbf{S}(j),\mathbf{Z}^{j-1})\\
&\quad + n(\tilde{R} - R_0 - R_K)\\
&= H(K_{j-1}|\mathcal{C},\mathbf{Z}^{j-1}) + n(\tilde{R} - R_0 - R_K).
\end{aligned}$$

From Proposition 1, $H(K_{j-1}|\mathcal{C},\mathbf{Z}^{j-1}) \ge n(R_K - \delta(\epsilon) - \delta'''(\epsilon))$, which implies that

$$\begin{aligned}
I(M_{j1};\mathbf{Z}(j)|\mathcal{C},M_{j0},\mathbf{S}(j),\mathbf{Z}^{j-1}) &\le n(\delta(\epsilon) + \delta'(\epsilon))\\
&\quad + n\delta'''(\epsilon).
\end{aligned}$$

This completes the analysis of the information leakage rate.

Rate Analysis: From the analysis of probability of error and information leakage rate, we see that the rate constraints are

$$\begin{aligned}
\tilde{R} &< I(U;Y,S) - \delta(\epsilon)\\
\tilde{R} - R_0 &> I(U;Z,S) + \delta(\epsilon)\\
R_K &< H(S|Z) - 4\delta(\epsilon)\\
R_0 + R_1 &\le \tilde{R}\\
R_1 &\le R_K\\
R &= R_0 + R_1\\
\tilde{R} &\ge 0, R_0 \ge 0, R_1 \ge 0, R_K \ge 0.
\end{aligned}$$

Using Fourier–Motzkin elimination (e.g., see Lecture 6 in [11]), we obtain

$$\begin{aligned}
R &< \max_{p(u),v(u,s),p(x|s,v)} \min\{I(U;Y,S) - I(U;Z,S)\\
&\qquad + H(S|Z), I(U;Y,S)\}\\
&\overset{(a)}{=} \max_{p(u),v(u,s),p(x|s,v)} \min\{I(V;Y|S) - I(V;Z|S)\\
&\qquad + H(S|Z), I(V;Y|S)\}
\end{aligned}$$

where (a) follows by the independence of U and S and the fact that V is a function of U and S.

Case 2: $R_{\mathrm{S-CSI-1}}$ With $I(U;Y,S) \le I(U;Z,S)$: In this case, only part of the key generated from the previous block is used to encrypt the message transmitted to the eavesdropper.

The other part of the key is used to generate additional uncertainty about the message sent at the eavesdropper to ensure that a sufficiently large secret key rate is achieved in the current block. Note that we only need to consider the case where $H(S|Z) - (I(U;Z,S) - I(U;Y,S)) > 0$.

Codebook Generation: Codebook generation again consists of two parts.

Message Codebook Generation: Let $\tilde{R} \geq R_d$ and $R \leq \tilde{R} - R_d$. Randomly and independently generate sequences $u^n(l)$, $l \in [1 : 2^{n\tilde{R}}]$, each according to $\prod_{i=1}^n p_U(u_i)$ and partition the set of indices $[1 : 2^{n\tilde{R}}]$ into 2^{nR_d} equal-size bins $\mathcal{C}(m_d)$, $m_d \in [1 : 2^{nR_d}]$. We further partition the set of indices in each bin $\mathcal{C}(m_d)$ into sub-bins, $\mathcal{C}(m_d, m)$, $m \in [1 : 2^{nR}]$.

Key Codebook Generation: We randomly bin the set of $s^n \in \mathcal{S}^n$ sequences into 2^{nR_K} bins $\mathcal{B}(k)$, $k \in [1 : 2^{nR_K}]$.

Encoding: We send $b-1$ messages over b n-transmission blocks. In the first block, we randomly select an index L. The encoder then computes $v_i = v(u_i(L), s_i)$, $i \in [1 : n]$, and transmits a randomly generates sequence X^n according to $\prod_{i=1}^n p_{X|S,V}(x_i|s_i, v_i)$. At the end of the first block, the encoder and decoder declare $k_1 \in [1 : 2^{nR_K}]$ such that $\mathbf{s}(1) \in \mathcal{B}(k_1)$ as the key to be used in block 2.

Encoding in block $j \in [2 : b]$ is as follows. We split the key k_{j-1} into two independent parts, $K_{j-1,d}$ and $K_{j-1,m}$ at rates R_d and R, respectively. To send message m_j, the encoder computes $m' = m_j \oplus k_{(j-1)m}$. This requires that $R_K \geq R + R_d$. The encoder then randomly selects an index $L \in \mathcal{C}(k_{(j-1)d}, m')$. At time $i \in [(j-1)n+1 : jn]$, it computes $v_i = (u_i(L), s_i)$, $i \in [1 : n]$, and transmits a randomly generated sequence X^n according to $\prod_{i=1}^n p_{X|S,V}(x_i|s_i, v_i)$.

Decoding and Analysis of the Probability of Error: At the end of block j, the decoder declares that $\hat{l}$ is sent if it is the unique index such that $(u^n(\hat{l}), \mathbf{Y}(j), \mathbf{S}(j)) \in \mathcal{T}_\epsilon^{(n)}$ and $\hat{l} \in \mathcal{C}(k_{(j-1)d})$. Otherwise, it declares an error. It then finds the index $\hat{m}'$ such that $u^n(\hat{l}) \in \mathcal{C}(k_{(j-1)d}, \hat{m}')$. Finally, it recovers $\hat{m}_j$ by computing $\hat{m}_j = (\hat{m}' - k_{(j-1)m})_{\mathrm{mod}\, 2^{nR}}$. Following similar steps to the analysis for Case 1, it can be shown that $P_e \to 0$ as $n \to \infty$ if $\tilde{R} - R_d < I(U;Y,S) - \delta(\epsilon)$.

Analysis of the Information Leakage Rate: Following the same steps as for Case 1, we can show that it suffices to upper bound the terms $I(M_j; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1})$ for $j \in [2 : b]$. Define $M_j' = M_j \oplus K_{(j-1)m}$ and consider

$$\begin{aligned}
&I(M_j; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}) \\
&\quad = H(M_j) - H(M_j|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^j) \\
&\quad \leq H(M_j) - H(M_j|\mathcal{C}, \mathbf{S}(j), K_{(j-1)d}, M_j', \mathbf{Z}^j) \\
&\quad = H(M_j) - H(M_j|\mathcal{C}, K_{(j-1)d}, M_j', \mathbf{Z}^{j-1}) \\
&\quad = H(M_j) - H(M_j|\mathcal{C}, \mathbf{Z}^{j-1}, K_{(j-1)d}) \\
&\qquad - H(M_j'|\mathcal{C}, \mathbf{Z}^{j-1}, K_{(j-1)d}, M_j) \\
&\qquad + H(M_j'|\mathcal{C}, \mathbf{Z}^{j-1}, K_{(j-1)d}) \\
&\quad = nR - H(M_j) + H(M_j'|\mathcal{C}, \mathbf{Z}^{j-1}, K_{(j-1)d}) \\
&\qquad - H(M_j'|\mathcal{C}, \mathbf{Z}^{j-1}, K_{(j-1)d}, M_j) \\
&\quad \leq nR - H(K_{(j-1)m}|\mathcal{C}, \mathbf{Z}^{j-1}, K_{(j-1)d}).
\end{aligned}$$

Next, we show that

$$I(K_{(j-1)m}; \mathbf{Z}^{j-1}|\mathcal{C}, K_{(j-1)d}) \leq n\delta'''(\epsilon) \tag{5}$$

$$H(K_{(j-1)m}|\mathcal{C}, K_{(j-1)d}) \geq n(R_K - R_d - \delta(\epsilon)) \tag{6}$$

which implies

$$\begin{aligned}
I(M_j; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}) &\leq nR - n(R_K - R_d) \\
&\quad + n(\delta'''(\epsilon) + \delta(\epsilon)).
\end{aligned}$$

Hence, the rate of information leakage tends to zero as $n \to \infty$ if $R \leq R_K - R_d$. To prove (5) and (6), we need the following.

Proposition 2: If $\tilde{R} > I(U;Z,S) + \delta(\epsilon)$ and $R_K < H(S|Z) - 4\delta(\epsilon)$, then for all $j \in [1 : b]$

1) $H(K_j|\mathcal{C}) \geq n(R_K - \delta(\epsilon))$;
2) $I(K_j; \mathbf{Z}(j)|\mathcal{C}) \leq n(\delta(\epsilon) + \delta'(\epsilon) + \delta''(\epsilon))$;
3) $I(K_j; \mathbf{Z}^j|\mathcal{C}) \leq n\delta'''(\epsilon)$, where $\delta'(\epsilon) \to 0$, $\delta''(\epsilon) \to 0$, and $\delta'''(\epsilon) \to 0$ as $\epsilon \to 0$.

The proof of this proposition is given in Appendix B.

Part 3 of Proposition 2 implies (5) since

$$\begin{aligned}
&I(K_{j-1}; \mathbf{Z}^{j-1}|\mathcal{C}) \\
&\quad = I(K_{(j-1)d}, K_{(j-1)m}; \mathbf{Z}^{j-1}|\mathcal{C}) \\
&\quad = I(K_{(j-1)d}; \mathbf{Z}^{j-1}|\mathcal{C}) + I(K_{(j-1)m}; \mathbf{Z}^{j-1}|\mathcal{C}, K_{(j-1)d}).
\end{aligned}$$

Part 1 of Proposition 2 implies (6) since $H(K_{(j-1)}|\mathcal{C}) = H(K_{(j-1)m}, K_{(j-1)d}|\mathcal{C}) \geq n(R_K - \delta(\epsilon))$, which implies that $H(K_{(j-1)m}|\mathcal{C}, K_{(j-1)d}) \geq n(R_K - R_d - \delta(\epsilon))$.

Rate Analysis: The following rate constraints are necessary for Case 2:

$$\begin{aligned}
\tilde{R} &> I(U;Z,S) + \delta(\epsilon) \\
\tilde{R} - R_d &< I(U;Y,S) - \delta(\epsilon) \\
R &\leq \tilde{R} - R_d \\
R_K &< H(S|Z) - 4\delta(\epsilon) \\
R &\leq R_K - R_d \\
R &\geq 0, R_K \geq 0, R_d \geq 0, \tilde{R} \geq 0.
\end{aligned}$$

Using Fourier–Motzkin elimination, we obtain

$$\begin{aligned}
R &< \max_{p(u),v(u,s),p(x|s,v)} \min \{I(U;Y,S) - I(U;Z,S) \\
&\qquad + H(S|Z), I(U;Y,S)\} \\
&= \max_{p(u),v(u,s),p(x|s,v)} \min \{I(V;Y|S) - I(V;Z|S) \\
&\qquad + H(S|Z), I(V;Y|S)\}.
\end{aligned}$$

Case 3: $R_{\mathrm{S-CSI-2}}$: Achievability of $R_{\mathrm{S-CSI-2}}$ uses the same techniques as Case 2 for $R_{\mathrm{S-CSI-1}}$. However, here the key generated in a block is used only to encrypt the message in the following block. The eavesdropper may be able to decode the message transmitted in a block, which would reduce the key rate generated at the end of that block. This is compensated for by the fact that the entire key is used for encryption. The

codebook generation, encoding, and analysis of probability of error and equivocation are, therefore, similar to that in Case 2.

As an outline, in each block, we generate a key K_j of size $2^{nH(S|V,Z)}$. In the next block, we encrypt the message M_{j+1} using K_j to obtain $M'_{j+1} = M_{j+1} \oplus K_j$. A codeword $V^n(M'_{j+1})$ is then selected from a codebook of size less than $2^{nI(V;Y,S)}$. This codeword is then transmitted by generating X_i according to $p_X(V_i, S_i)$ for $i \in [1 : n]$. The decoder first decodes M'_{j+1} using $(\mathbf{Y}(j+1), \mathbf{S}(j+1))$ and then decrypts the message using $M_{j+1} = (M'_{j+1} - K_j)_{\mathrm{mod}\ 2^{nR_K}}$.

Remark 4.1: An important difference between the schemes for achieving $R_{\mathrm{S-CSI-1}}$ and $R_{\mathrm{S-CSI-2}}$ is that the transmitted codeword in the scheme for $R_{\mathrm{S-CSI-1}}$ can be kept secret from the eavesdropper through a combination of Wyner wiretap coding and encryption of the codebook using the secret key. This fact was made clear in Case 2 of the proof, where part of the key was explicitly used to encrypt the codebook instead of the message. On the other hand, no effort was made to keep the transmitted codeword secret from the eavesdropper in the scheme for $R_{\mathrm{S-CSI-2}}$. Hence, one cannot assume that the eavesdropper has no information about the codeword sent. As such, we did not use the Shannon strategy as in Fig. 1 in the encoding part for $R_{\mathrm{S-CSI-2}}$. If we had used the Shannon strategy, we would have obtained the expression

$$R'_{\mathrm{S-CSI-2}} = \max_{p(u),v(u,s),p(x|s,v)} \{H(S|Z,U), I(U;Y,S)\} = \max_{p(u),v(u,s),p(x|s,v)} \{H(S|Z,U), I(V;Y|S)\}.$$

The first term in $R'_{\mathrm{S-CSI-2}}$, $H(S|Z,U)$, is the rate of the secret key rate generated assuming that the eavesdropper knows the transmitted codeword. As a result of this term, $R'_{\mathrm{S-CSI-2}}$ is simply a special case of our original $R_{\mathrm{S-CSI-2}}$ expression, where we maximized the rate over $p(v)$ (V independent of S). This is in contrast to the equivalent characterization of $R_{\mathrm{S-CSI-1}}$ in (3), where we were able to perform the maximization of the rate over $p(v|s)$.

We now turn to the proof of achievability of $R_{\mathrm{S-CSI-2}}$.

Codebook Generation: Codebook generation again consists of two parts.

Message Codebook Generation: Randomly and independently generate sequences $v^n(l)$, $l \in [1 : 2^{nR}]$, each according to $\prod_{i=1}^n p_V(v_i)$.

Key Codebook Generation: Set $R_K = R$. Randomly bin the set of $s^n \in \mathcal{S}^n$ sequences into 2^{nR_K} bins $\mathcal{B}(k)$, $k \in [1 : 2^{nR_K}]$.

Encoding: We send $b-1$ messages over b n-transmission blocks. In the first block, we randomly select an index $L \in [1 : 2^{nR}]$. The encoder then selects $v^n(L)$ and transmits a randomly generated sequence X^n according to $\prod_{i=1}^n p_{X|S,V}(x_i|s_i, v_i)$. At the end of the first block, the encoder and the decoder declare the index $k_1 \in [1 : 2^{nR_K}]$ such that $\mathbf{s}(1) \in \mathcal{B}(k_1)$ as the key to be used in block 2.

Encoding in block $j \in [2 : b]$ is as follows. To send message m_j, the encoder computes the encrypted message $m' = m_j \oplus k_{j-1}$. It then selects the sequence $v^n(m')$. At time $i \in [(j-1)n + 1 : jn]$, it transmits a randomly generated sequence X^n according to $\prod_{i=1}^n p_{X|S,V}(x_i|s_i, v_i)$.

Decoding and Analysis of the Probability of Error: At the end of block j, the decoder declares that $\hat{m}'$ is sent if it is the unique index such that $(v^n(\hat{m}'), \mathbf{Y}(j), \mathbf{S}(j)) \in \mathcal{T}_\epsilon^{(n)}$. Otherwise, it declares an error. It then recovers $\hat{m}_j$ by computing $\hat{m}_j = (\hat{m}' - k_{j-1})_{\mathrm{mod}\ 2^{nR_K}}$. Following similar steps to the analysis for Case 1, it can be shown that the probability of error tends to zero as $n \to \infty$ if $R < I(V;Y,S) - \delta(\epsilon)$.

Analysis of the Information Leakage Rate: Following the same steps as for Case 1, we can show that it suffices to upper bound the terms $I(M_j; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1})$ for $j \in [2 : b]$. Consider

$$\begin{aligned}
&I(M_j; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}) \\
&\quad = H(M_j) - H(M_j|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^j) \\
&\quad \le H(M_j) - H(M_j|\mathcal{C}, \mathbf{S}(j), M_j \oplus K_{j-1}, \mathbf{Z}^j) \\
&\quad = H(M_j) - H(M_j|\mathcal{C}, M_j \oplus K_{j-1}, \mathbf{Z}^{j-1}) \\
&\quad = H(M_j) - H(M_j \oplus K_{j-1}, M_j|\mathcal{C}, \mathbf{Z}^{j-1}) \\
&\qquad + H(M_j \oplus K_{j-1}|\mathcal{C}, \mathbf{Z}^{j-1}) \\
&\quad \le nR - H(M_j|\mathcal{C}, \mathbf{Z}^{j-1}) \\
&\qquad - H(M_j \oplus K_{j-1}|\mathcal{C}, \mathbf{Z}^{j-1}, M_j) + nR \\
&\quad = nR - H(K_{j-1}|\mathcal{C}, \mathbf{Z}^{j-1}).
\end{aligned}$$

Next, we show that

$$I(K_{j-1}; \mathbf{Z}^{j-1}|\mathcal{C}) \le n\delta''(\epsilon) \tag{7}$$

$$H(K_{j-1}|\mathcal{C}) \ge n(R_K - \delta(\epsilon)). \tag{8}$$

which would imply

$$I(M_j; \mathbf{Z}(j)|\mathcal{C}, \mathbf{S}(j), \mathbf{Z}^{j-1}) \le n(\delta''(\epsilon) + \delta(\epsilon)).$$

To prove (7) and (8), we will use the following.

Proposition 3: If $R_K < H(S|Z,V) - 4\delta(\epsilon)$, then for all $j \in [1 : b]$

1) $H(K_j|\mathcal{C}) \ge n(R_K - \delta(\epsilon))$;
2) $I(K_j; \mathbf{Z}(j)|\mathcal{C}) \le n\delta'(\epsilon)$;
3) $I(K_j; \mathbf{Z}^j|\mathcal{C}) \le n\delta''(\epsilon)$, where $\delta'(\epsilon) \to 0$ and $\delta''(\epsilon) \to 0$ as $\epsilon \to 0$.

The proof of this proposition is given in Appendix C. It is clear that (7) and (8) are implied by Proposition 3, which completes the analysis of information leakage rate.

Rate Analysis: The following rate constraints are necessary for Case 3:

$$\begin{aligned}
R &= R_K \\
R &< I(V;Y,S) - \delta(\epsilon) \\
R_K &< H(S|Z,V) - 4\delta(\epsilon) \\
R_K &\ge 0, R \ge 0.
\end{aligned}$$

Using Fourier–Motzkin elimination, these constraints imply the achievability of

$$R < \max_{p(v)p(x|s,v)} \min\{H(S|Z,V), I(V;Y|S)\}.$$

V. Proof of Theorem 2

For any sequence of codes with probability of error and leakage rate that approach zero as $n \to \infty$, consider

$$\begin{aligned}
nR &= H(M) \\
&\overset{(a)}{\leq} I(M;Y^n,S^n) + n\epsilon_n \\
&\overset{(b)}{\leq} I(M;Y^n,S^n) - I(M;Z^n) + 2n\epsilon_n \\
&= \sum_{i=1}^{n} \big(I(M;Y_i,S_i|Y_{i+1}^n,S_{i+1}^n) \\
&\qquad - I(M;Z_i|Z^{i-1})\big) + 2n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} \big(I(M,Z^{i-1};Y_i,S_i|Y_{i+1}^n,S_{i+1}^n) \\
&\qquad - I(M,Y_{i+1}^n,S_{i+1}^n;Z_i|Z^{i-1})\big) + 2n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} \big(I(M;Y_i,S_i|Y_{i+1}^n,S_{i+1}^n,Z^{i-1}) \\
&\qquad - I(M;Z_i|Y_{i+1}^n,S_{i+1}^n,Z^{i-1})\big) + 2n\epsilon_n \\
&\overset{(e)}{=} \sum_{i=1}^{n} \big(I(V_{1i};Y_i,S_i|U_i) - I(V_{1i};Z_i|U_i)\big) + 2n\epsilon_n \\
&= \sum_{i=1}^{n} \big(I(V_{1i};Y_i,S_i|U_i) - I(V_{1i};Z_i,S_i|U_i) \\
&\qquad + I(V_{1i};S_i|Z_i,U_i)\big) + 2n\epsilon_n \\
&\leq \sum_{i=1}^{n} \big(I(V_{1i};Y_i,S_i|U_i) - I(V_{1i};Z_i,S_i|U_i) \\
&\qquad + H(S_i|Z_i,U_i)\big) + 2n\epsilon_n \\
&\leq \sum_{i=1}^{n} \big(I(V_{1i};Y_i|U_i,S_i) - I(V_{1i};Z_i,S_i|U_i,S_i) \\
&\qquad + H(S_i|Z_i,U_i)\big) + 2n\epsilon_n \\
&\overset{(f)}{=} n\big(I(V_1;Y|U,S) - I(V_1;Z|U,S)\big) \\
&\quad + nH(S|Z,U) + 2n\epsilon_n
\end{aligned}$$

where (a) follows by Fano's inequality; (b) follows by the secrecy condition; (c) and (d) follow by the Csiszár sum identity; (e) follows by the identifications $U_i = (Y_{i+1}^n, S_{i+1}^n, Z^{i-1})$ and $V_{1i} = (M, Y_{i+1}^n, S_{i+1}^n, Z^{i-1})$; and (f) follows by using the time-sharing random variable Q and defining $U = (U_Q, Q)$, $V_1 = (V_{1Q}, Q)$, $S = S_Q$, $Y = Y_Q$, and $Z = Z_Q$.

For the second upper bound, we have

$$\begin{aligned}
nR &\leq I(M;Y^n,S^n) + n\epsilon_n \\
&\overset{(a)}{=} I(M;Y^n|S^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(M;Y_i|S^n,Y_{i+1}^n) \\
&\leq \sum_{i=1}^{n} I(M,Y_{i+1}^n,Z^{i-1},S_{i+1}^n,S^{i-1};Y_i|S_i) \\
&\overset{(b)}{=} \sum_{i=1}^{n} I(V_{2i};Y_i|S_i) \\
&= nI(V_{2Q};Y|S,Q) \\
&\overset{(c)}{\leq} nI(V_2;Y|S)
\end{aligned}$$

where (a) follows by the independence of M and S^n; (b) follows by the identification $V_{2i} = (M, Y_{i+1}^n, Z^{i-1}, S_{i+1}^n, S^{i-1})$; and (c) follows by defining $V_2 \doteq (V_{2Q}, Q)$. The bounds on cardinality of the auxiliary random variables follow from standard techniques (e.g., see [11, App. C]).

VI. Conclusion

We established bounds on the secrecy capacity of the wiretap channel with state information causally available at the encoder and the decoder. We showed that our lower bound can be strictly larger than the best known lower bound for the noncausal state information case. The upper bound holds when the state information is available noncausally at the encoder and the decoder. We showed that the bounds are tight for several classes of wiretap channels with state.

As we have seen, the secrecy capacity for several special classes of the wiretap channels with state available at both the encoder and the decoder does not depend on whether the state is available causally or noncausally. An interesting question to explore is whether this observation holds in general for our setup.

We used key generation from state information to improve the message transmission rate. It may be possible to extend this idea to the case when the state information is available only at the encoder. This case, however, is not straightforward to analyze since it would be necessary for the encoder to reveal some state information to the decoder (and, hence, partially to the eavesdropper) in order to agree on a secret key, which would reduce the wiretap coding part of the rate.

Appendix A
Proof of Proposition 1

1) The proof of this part follows largely from Lemma 2 in [11, Lecture 22]. For completeness, we give the proof here. Consider

$$\begin{aligned}
H(K_j|\mathcal{C}) &\geq P\{S^n \in \mathcal{T}_\epsilon^{(n)}\} H(K_j|\mathcal{C}, \mathbf{S}(j) \in \mathcal{T}_\epsilon^{(n)}) \\
&\geq (1-\epsilon_n') H(K_j|\mathcal{C}, \mathbf{S}(j) \in \mathcal{T}_\epsilon^{(n)}).
\end{aligned}$$

Let $P(k_j)$ be the *random* pmf of K_j given $\{\mathbf{S}(j) \in \mathcal{T}_\epsilon^{(n)}\}$, where the randomness is induced by the random bin assignment (codebook) $\mathcal{C}$.

By symmetry, $P(k_j)$, $k_j \in [1 : 2^{nR_K}]$, are identically distributed. We express $P(1)$ in terms of a weighted sum of indicator functions as

$$P(1) = \sum_{s^n \in \mathcal{T}_\epsilon^{(n)}} \frac{p(s^n)}{P\{S^n \in \mathcal{T}_\epsilon^{(n)}\}} \cdot I_{\{s^n \in \mathcal{B}(1)\}}.$$

It can be easily shown that

$$\begin{aligned}
&\mathrm{E}_{\mathcal{C}}(P(1)) = 2^{-nR_K} \\
&\mathrm{Var}(P(1)) \\
&= 2^{-nR_K}(1 - 2^{-nR_K}) \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}} \left(\frac{p(s^n)}{P\{\mathbf{S}(j) \in \mathcal{T}_\epsilon^{(n)}\}}\right)^2 \\
&\leq 2^{-nR_K} 2^{n(H(S)+\delta(\epsilon))} \frac{2^{-2n(H(S)-\delta(\epsilon))}}{(1-\epsilon_n')^2} \\
&\leq 2^{-n(R_K + H(S) - 4\delta(\epsilon))}
\end{aligned}$$

for sufficiently large n.
By the Chebyshev inequality

$$\begin{aligned}&P\{|P(1)-\mathrm{E}(P(1))|\geq \epsilon \mathrm{E}(P(1))\}\\&\leq \frac{\mathrm{Var}(P(1))}{(\epsilon \mathrm{E}(P(1)))^2}\\&\leq \frac{2^{-n(H(S)-R_K-4\delta(\epsilon))}}{\epsilon^2}.\end{aligned}$$

Note that if $R_K < H(S)-4\delta(\epsilon)$, this probability $\to 0$ as $n\to\infty$. Now, by symmetry

$$\begin{aligned}&H(K_1|\mathcal{C},\mathbf{S}(j)\in\mathcal{T}_\epsilon^{(n)})\\&=2^{nR_K}\mathrm{E}(P(1))\log(1/P(1)))\\&\geq 2^{nR_K}P\{|P(1)-\mathrm{E}(P(1))|<\epsilon 2^{-nR_K}\}.\\&\quad \mathrm{E}\big(P(1)\log(1/P(1))\,\big|\,|P(1)-\mathrm{E}(P(1))|<\epsilon 2^{-nR_K}\big)\\&\geq \left(1-\frac{2^{-n(H(S)-R_K-4\delta(\epsilon))}}{\epsilon^2}\right).\\&\quad (nR_K(1-\epsilon)-(1-\epsilon)\log(1+\epsilon))\\&\geq n(R_K-\delta(\epsilon))\end{aligned}$$

for sufficiently large n and $R_K < H(S)-4\delta(\epsilon)$.
Thus, we have shown that if $R_K < H(S)-4\delta(\epsilon)$, $H(K_j|\mathcal{C})\geq n(R_K-\delta(\epsilon))$ for n sufficiently large. This completes the proof of part 1 of Proposition 1. Note now that since $H(S|Z)\leq H(S)$, the same results also holds if $R_K \leq H(S|Z)-4\delta(\epsilon)$.

2) We need to show that if $R_K < H(S|Z)-3\delta(\epsilon)$, then $I(K_j;\mathbf{Z}(j)|\mathcal{C})\leq 2n\delta(\epsilon)$ for every $j\in[1:b]$. We have

$$I(K_j;\mathbf{Z}(j)|\mathcal{C})=I(\mathbf{S}(j);\mathbf{Z}(j)|\mathcal{C})-I(\mathbf{S}(j);\mathbf{Z}(j)|K_j,\mathcal{C})$$

We analyze the terms separately. For the first term, we have

$$\begin{aligned}&I(\mathbf{S}(j);\mathbf{Z}(j)|\mathcal{C})\\&=I(\mathbf{S}(j),L;\mathbf{Z}(j)|\mathcal{C})-I(L;\mathbf{Z}(j)|\mathbf{S}(j),\mathcal{C})\\&\leq I(U^n,\mathbf{S}(j);\mathbf{Z}(j)|\mathcal{C})-H(L|\mathbf{S}(j),\mathcal{C})\\&\quad +H(L|\mathcal{C},\mathbf{S}(j),\mathbf{Z}(j))\\&\leq nI(U,S;Z)-H(L|\mathbf{S}(j),\mathcal{C})+H(L|\mathcal{C},\mathbf{S}(j),Z^n)\\&\overset{(a)}{\leq} nI(U,S;Z)-H(L|\mathbf{S}(j),\mathcal{C})\\&\quad +n(\tilde{R}-I(U;Z,S)+\delta'(\epsilon))\\&=n\tilde{R}-H(M_{j0}|\mathcal{C})-H(M_{j1}\oplus K_{j-1}|\mathcal{C},M_{j0})\\&\quad -H(L|M_{j0},M_{j1}\oplus K_{j-1},\mathcal{C})+nI(S;Z)+n\delta'(\epsilon)\\&\leq n\tilde{R}-nR_0-H(M_{j1}\oplus K_{j-1}|\mathcal{C},M_{j0},K_{j-1})\\&\quad -n(\tilde{R}-R_0-R_K)+nI(S;Z)+n\delta'(\epsilon)\\&=nR_K-H(M_{j1}|\mathcal{C},M_{j0},K_{j-1})+nI(S;Z)+n\delta'(\epsilon)\\&=n(I(S;Z)+\delta'(\epsilon))\end{aligned}$$

where step (a) follows by applying Lemma 1, which holds since $\tilde{R}-R_0 > I(U;Z,S)+\delta(\epsilon)$ and $P((U^n(L),\mathbf{S}(j),\mathbf{Z}(j))\in\mathcal{T}_\epsilon^{(n)})\to 1$ as $n\to\infty$. For the second term, we have

$$\begin{aligned}&I(\mathbf{S}(j);\mathbf{Z}(j)|K_j,\mathcal{C})\\&=H(\mathbf{S}(j)|K_j,\mathcal{C})-H(\mathbf{S}(j)|\mathbf{Z}(j),K_j,\mathcal{C})\\&=H(\mathbf{S}(j),K_j|\mathcal{C})-H(K_j|\mathcal{C})-H(\mathbf{S}(j)|\mathbf{Z}(j),K_j,\mathcal{C})\\&\geq nH(S)-nR_K-H(\mathbf{S}(j)|\mathbf{Z}(j),K_j,\mathcal{C})\\&\geq n(H(S)-R_K)-H(\mathbf{S}(j)|\mathbf{Z}(j),K_j)\\&\overset{(b)}{\geq} n(H(S)-R_K)-n(H(S|Z)-R_K+\delta''(\epsilon))\\&=nI(S;Z)-n\delta'(\epsilon)\end{aligned}$$

where (b) follows by showing that $H(\mathbf{S}(j)|\mathbf{Z}(j),K_j)\leq n(H(S|Z)-R_K+\delta(\epsilon))$. This requires the condition $R_K < H(S|Z)-3\delta(\epsilon)$. Combining the bounds for the two expressions gives $I(K_j;\mathbf{Z}(j)|\mathcal{C})\leq n(\delta'(\epsilon)+\delta''(\epsilon))$.

Proof of Step (b): Give an arbitrary ordering to the set of all state sequences s^n with $\mathbf{S}(j)=s^n(T)$ for some $T\in[1:2^{n\log|\mathcal{S}|}]$. Hence, $H(\mathbf{S}(j)|\mathbf{Z}(j),K)=H(T|K,\mathbf{Z}(j))$.

From the coding scheme, we know that $P\{(s^n(T),\mathbf{Z}(j))\in\mathcal{T}_\epsilon^{(n)}\}\to 1$ as $n\to\infty$. Note here that T is random and corresponds to the realization of S^n.

Now, fix $T=t,\mathbf{Z}(j)=z^n,K=k$, and define $N(z^n,k,t)=|\tilde{l}\in[1:|\mathcal{T}_\epsilon^{(n)}(S)|]:(s^n(\tilde{l}),z^n)\in\mathcal{T}_\epsilon^{(n)},\ \tilde{l}\neq t,\ s^n(\tilde{l})\in\mathcal{B}(k)|$. For $z^n\notin\mathcal{T}_\epsilon^{(n)}$, $N(z^n,k,t)=0$. For $z^n\in\mathcal{T}_\epsilon^{(n)}$, it is easy to show that

$$\frac{|\mathcal{T}_\epsilon^{(n)}(S|Z)|-1}{2^{nR_K}}\leq \mathrm{E}(N(z^n,k,t))\leq \frac{|\mathcal{T}_\epsilon^{(n)}(S|Z)|}{2^{nR_K}}$$

$$\mathrm{Var}(N(z^n,k,t))\leq \frac{|\mathcal{T}_\epsilon^{(n)}(S|Z)|}{2^{nR_K}}.$$

By the Chebyshev inequality

$$\begin{aligned}&P\{N(z^n,k,t)\geq(1+\epsilon)\mathrm{E}(N(z^n,k,t))\}\\&\leq\frac{\mathrm{Var}(N(z^n,k,t))}{(\epsilon\mathrm{E}(N(z^n,k,t)))^2}\\&\leq\frac{2^{-n(H(S|Z)-3\delta(\epsilon)-R_K)}}{\epsilon^2}.\end{aligned}$$

Note that $P\{N(z^n,k,t)\geq(1+\epsilon)\mathrm{E}(N(z^n,k,t))\}\to 0$ as $n\to\infty$ if $R < H(S|Z)-3\delta(\epsilon)$. Now, define the following events:

$$\begin{aligned}\mathcal{E}_1&=\{(\mathbf{S}(j),\mathbf{Z}(j))\notin\mathcal{T}_\epsilon^{(n)}\}\\\mathcal{E}_2&=\{N(\mathbf{Z}(j),K,T)\geq(1+\epsilon)\mathrm{E}(N(\mathbf{Z}(j),K,T))\}.\end{aligned}$$

Let $E=0$ if $\mathcal{E}_1^c\cap\mathcal{E}_2^c$ occurs and 1 otherwise. We have

$$\begin{aligned}&P(E=1)\\&\leq P(\mathcal{E}_1)+P(\mathcal{E}_2)\\&\leq\sum_{(z^n,s^n(t))\in\mathcal{T}_\epsilon^{(n)},\,k}(p(z^n,t,k).\\&\qquad P\{N(z^n,k,t)\geq(1+\epsilon)\mathrm{E}(N(z^n,k,t))\})\\&\quad +P(\mathcal{E}_1)\\&\quad +P\{(s^n(T),\mathbf{Z}(j))\notin\mathcal{T}_\epsilon^{(n)}\}.\end{aligned}$$

$P\{(s^n(T),\mathbf{Z}(j))\notin\mathcal{T}_\epsilon^{(n)}\}=P(\mathcal{E}_1)$ and $P(\mathcal{E}_1)\to 0$ as $n\to\infty$ by the coding scheme. For the second term, $P\{N(z^n,k,t)\geq$

$(1+\epsilon)\mathrm{E}(N(z^n,k,t))\} \to 0$ as $n \to \infty$ if $R < H(S|Z)-3\delta(\epsilon)$. Hence, $P(E=1) \to 0$ as $n \to \infty$ if $R < H(S|Z) - 3\delta(\epsilon)$.

We can now bound $H(T|K,Z^n)$ by

$$\begin{aligned} H(T|K,Z^n) &\le 1 + P(E=1)H(T|K,Z^n,E=1) \\ &\quad + H(T|K,Z^n,E=0) \\ &\le n(H(S|Z) - R_K + \delta(\epsilon)). \end{aligned}$$

This completes the proof of part 2.

1) To upper bound $I(K_j;\mathbf{Z}^j|\mathcal{C})$, we use an induction argument assuming that $I(K_{j-1};\mathbf{Z}^{j-1}|\mathcal{C}) \le n\delta_{j-1}(\epsilon)$, where $\delta_{j-1}(\epsilon) \to 0$ as $\epsilon \to 0$. Note that the proof for $j=2$ follows from part 2. Consider

$$\begin{aligned} &I(K_j;\mathbf{Z}^j|\mathcal{C}) \\ &= I(K_j;\mathbf{Z}(j)|\mathcal{C}) + I(K_j;\mathbf{Z}^{j-1}|\mathcal{C},\mathbf{Z}(j)) \\ &\overset{(a)}{\le} n(\delta'(\epsilon)+\delta''(\epsilon)) + I(K_j;\mathbf{Z}^{j-1}|\mathcal{C},\mathbf{Z}(j)) \\ &= H(\mathbf{Z}^{j-1}|\mathcal{C},\mathbf{Z}(j)) - H(\mathbf{Z}^{j-1}|\mathcal{C},\mathbf{Z}(j),K_j) \\ &\quad + n(\delta'(\epsilon)+\delta''(\epsilon)) \\ &\le H(\mathbf{Z}^{j-1}|\mathcal{C}) - H(\mathbf{Z}^{j-1}|\mathcal{C},K_{j-1},\mathbf{Z}(j),K_j) \\ &\quad + n(\delta'(\epsilon)+\delta''(\epsilon)) \\ &\overset{(b)}{=} H(\mathbf{Z}^{j-1}|\mathcal{C}) - H(\mathbf{Z}^{j-1}|\mathcal{C},K_{j-1}) \\ &\quad + n(\delta'(\epsilon)+\delta''(\epsilon)) \\ &= I(K_{j-1};\mathbf{Z}^{j-1}|\mathcal{C}) + n(\delta'(\epsilon)+\delta''(\epsilon)) \\ &\overset{(c)}{\le} n\delta_{j-1}(\epsilon) + n(\delta'(\epsilon)+\delta''(\epsilon)) \end{aligned}$$

where (a) follows from part 2 of the Proposition; (b) follows from the Markov Chain relation $\mathbf{Z}^{j-1} \to K_{j-1} \to (\mathbf{Z}(j),K_j)$; and (c) follows from the induction hypothesis. This completes the proof since the last line implies that there exists a $\delta'''(\epsilon)$, where $\delta'''(\epsilon) \to 0$ as $\epsilon \to 0$, that upper bounds $I(K_j;\mathbf{Z}^j|\mathcal{C})$ for $j \in [1:b]$.

Appendix B
Proof of Proposition 2

1) We first show that if $R_K < H(S) - 4\delta(\epsilon)$, then $H(K_j|\mathcal{C}) \ge n(R_K - \delta(\epsilon))$. This is done in the same manner as for part 1 of Proposition 1. The proof is, therefore, omitted.
2) We need to show that if $R_K < H(S|Z) - 3\delta(\epsilon)$, then $I(K_j;\mathbf{Z}(j)|\mathcal{C}) \le 2n\delta(\epsilon)$ for every $j \in [1:b]$. We have

$$I(K_j;\mathbf{Z}(j)|\mathcal{C}) = I(\mathbf{S}(j);\mathbf{Z}(j)|\mathcal{C}) - I(\mathbf{S}(j);\mathbf{Z}(j)|K_j,\mathcal{C}).$$

We analyze the terms separately. For the first term, we have

$$\begin{aligned} &I(\mathbf{S}(j);\mathbf{Z}(j)|\mathcal{C}) \\ &= I(\mathbf{S}(j),L;\mathbf{Z}(j)|\mathcal{C}) - I(L;\mathbf{Z}(j)|\mathbf{S}(j),\mathcal{C}) \\ &\le I(U^n,\mathbf{S}(j);\mathbf{Z}(j)|\mathcal{C}) - H(L|\mathbf{S}(j),\mathcal{C}) \\ &\quad + H(L|\mathcal{C},\mathbf{S}(j),\mathbf{Z}(j)) \\ &\le nI(U,S;Z) - H(L|\mathbf{S}(j),\mathcal{C}) \\ &\quad + H(L|\mathcal{C},\mathbf{S}(j),Z^n) \\ &\overset{(a)}{\le} nI(U,S;Z) - H(L|\mathcal{C}) \\ &\quad + n(\tilde{R} - I(U;Z,S) + \delta'(\epsilon)) \\ &= n\tilde{R} - H(K_{(j-1)d}|\mathcal{C}) \\ &\quad - H(K_{(j-1)m} \oplus M_j|\mathcal{C}) + nI(S;Z) \\ &\quad - H(L|K_{(j-1)m} \oplus M_j, K_{(j-1)d}) + n\delta'(\epsilon) \\ &\overset{(b)}{\le} n(\tilde{R} - R_d - R - \tilde{R} + R_d + R + \delta(\epsilon) \\ &\quad + \delta'(\epsilon)) + nI(S;Z) \\ &= n(I(S;Z) + \delta(\epsilon) + \delta'(\epsilon)) \end{aligned}$$

where step (a) follows by applying Lemma 1, which holds by the condition $\tilde{R} > I(U;Z,S) + \delta(\epsilon)$ and the fact that $P((U^n(L),\mathbf{S}(j),\mathbf{Z}(j)) \in \mathcal{T}_\epsilon^{(n)}) \to 1$ as $n \to \infty$ from the encoding scheme. Step (b) follows from part 1 of Proposition 2: $H(K_{j-1}|\mathcal{C}) \ge n(R_K - \delta(\epsilon))$, which implies that $H(K_{(j-1)d}|\mathcal{C}) \ge n(R_d - \delta(\epsilon))$. Note that we implicitly assumed $j \ge 2$. The case of $j=1$ is straightforward since $H(L|\mathcal{C}) = n\tilde{R}$ by the fact that we transmit a codeword picked uniformly at random.

The proof that $I(\mathbf{S}(j);\mathbf{Z}(j)|K_j,\mathcal{C}) \ge nI(S;Z) - n\delta''(\epsilon)$ follows the same steps as the proof of part 2 of Proposition 1 and requires the same condition that $R_K < H(S|Z) - 3\delta(\epsilon)$.

3) This part is proved in the same manner as part 3 of Proposition 1.

Appendix C
Proof of Proposition 3

1) We first show that if $R_K < H(S) - 4\delta(\epsilon)$, then $H(K_j|\mathcal{C}) \ge n(R_K - \delta(\epsilon))$. This is done in the same manner as part 1 of Proposition 1. The proof is, therefore, omitted.
2) We need to show that if $R_K < H(S|Z,V) - 3\delta(\epsilon)$, then $I(K_j;\mathbf{Z}(j)|\mathcal{C}) \le n\delta(\epsilon)$ for every $j \in [1:b]$. Consider

$$\begin{aligned} I(K_j;\mathbf{Z}(j)|\mathcal{C}) &\le I(K_j;\mathbf{Z}(j),U^n|\mathcal{C}) \\ &= I(\mathbf{S}(j);\mathbf{Z}(j),U^n|\mathcal{C}) \\ &\quad - I(\mathbf{S}(j);\mathbf{Z}(j),U^n|K_j,\mathcal{C}). \end{aligned}$$

We analyze each term separately. For the first term, we have

$$\begin{aligned} &I(\mathbf{S}(j);\mathbf{Z}(j),V^n|\mathcal{C}) \\ &= I(\mathbf{S}(j);\mathbf{Z}(j)|V^n,\mathcal{C}) \\ &= \sum_{i=1}^n (H(\mathbf{Z}_i(j)|\mathcal{C},V^n,\mathbf{Z}^{i-1}(j)) \\ &\qquad - H(\mathbf{Z}_i(j)|\mathcal{C},V^n,\mathbf{S}(j),\mathbf{Z}^{i-1}(j))) \\ &\le \sum_{i=1}^n (H(\mathbf{Z}_i(j)|\mathcal{C},V_i) - H(\mathbf{Z}_i(j)|\mathcal{C},V_i,\mathbf{S}_i(j))) \\ &\le n(H(Z|V) - H(Z|V,S)) \\ &= nI(Z;S|V) = nI(Z,V;S). \end{aligned}$$

For the second term, we have

$$\begin{aligned}
&I(\mathbf{S}(j); \mathbf{Z}(j), V^n | K_j, \mathcal{C}) \\
&= H(\mathbf{S}(j)|K_j, \mathcal{C}) - H(\mathbf{S}(j)|\mathbf{Z}(j), V^n, K_j, \mathcal{C}) \\
&= H(\mathbf{S}(j), K_j|\mathcal{C}) - H(K_j|\mathcal{C}) \\
&\quad - H(\mathbf{S}(j)|\mathbf{Z}(j), V^n, K_j, \mathcal{C}) \\
&\geq nH(S) - nR_K - H(\mathbf{S}(j)|\mathbf{Z}(j), V^n, K_j, \mathcal{C}) \\
&\geq n(H(S) - R_K) - H(\mathbf{S}(j)|\mathbf{Z}(j), V^n, K_j) \\
&\overset{(b)}{\geq} n(H(S) - R_K) - n(H(S|Z, V) - R_K + \delta'(\epsilon)) \\
&= nI(S; Z, V) - n\delta'(\epsilon).
\end{aligned}$$

The proof of step (b) follows the same steps as in the proof of part 2 of Proposition 1. We can show that step (b) holds if $R_K < H(S|Z, V) - 3\delta(\epsilon)$.

Combining the two terms then give the required upper bound which completes the proof of Part 2.

3) This part is proved in the same manner as part 3 of Proposition 1.

ACKNOWLEDGMENT

The authors would like to thank both the anonymous reviewers for this paper and for the conference version of this paper for helpful comments and questions that helped improve the paper.

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[3] H. Yamamoto, "Rate-distortion theory for the Shannon cipher system," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 827–835, May 1997.

[4] Y. Chen and A. J. Han Vinck, "Wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 395–402, Jan. 2006.

[5] W. Liu and B. Chen, "Wiretap channel with two-sided state information," in *Proc. 41st Asilomar Conf. Signals, Syst. Comp.*, Pacific Grove, CA, Nov. 2007, pp. 893–897.

[6] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, Jul. 2009, pp. 2286–2290.

[7] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[8] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography—I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[9] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.

[10] E. Ardestanizadeh, M. Franceschetti, T. Javidi, and Y. H. Kim, "Wiretap channel with rate-limited feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5353–5361, Dec. 2009.

[11] A. El Gamal and Y. H. Kim, *Network Information Theory*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[12] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[13] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Information Theory (Second Colloq., Keszthely, 1975)*. Amsterdam, The Netherlands: North Holland, 1977, pp. 411–423.

[14] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

[15] Y. K. Chia and A. El Gamal, "3-receiver broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, to be published.

Yeow-Khiang Chia received his M.Eng. degree in Electrical Engineering from Imperial College, London. He received his M.Sc. and Ph.D. degrees, both in Electrical Engineering, from Stanford University in 2011. He is currently working as a scientist with the Institute for Infocomm Research, Singapore. He was a recipient of the Stanford Graduate Fellowship (SGF) and the National Science Scholarship (NSS) from the Agency for Science, Technology and Research, Singapore (A*STAR) for his Ph.D. studies.

Abbas El Gamal (S'71–M'73–SM'83–F'00) received the B.Sc. (honors) degree in electrical engineering from Cairo University in 1972 and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1977 and 1978, respectively.

From 1978 to 1980, he was an Assistant Professor in the Department of Electrical Engineering at the University of Southern California (USC). He has been on the Stanford faculty since 1981, where he is currently the Hitachi America Professor in the School of Engineering and the Director of the Information Systems lab in the department of electrical engineering. His research interest and contributions are in the areas of network information theory, wireless communications, digital imaging, and integrated circuit design. He has authored or coauthored over 200 papers and 30 patents in these areas. He is coauthor of the book *Network Information Theory* (Cambridge Press 2011).

Dr. El Gamal has won several honors and awards, including the 2004 Infocom best paper award, the 2009 Padovani lecture, and the 2012 Shannon Award. He has been serving on the Board of Governors of the IEEE IT Society since 2009 and is currently the Second Vice President.