Strong Secrecy of Arbitrarily Varying Wiretap Channel With Constraints

Yiqi Chen, Dan He, Chenhao Ying, and Yuan Luo, *Member, IEEE*

***Abstract*— The strong secrecy transmission problem of the arbitrarily varying wiretap channel (AVWC) with input and state constraints is investigated in this paper. First, a stochastic-encoder code lower bound of the strong secrecy capacity is established by applying the type argument and Csiszár's almost independent coloring lemma. Then, a superposition stochastic-encoder code lower bound of the secrecy capacity is provided. The superposition stochastic-encoder code lower bound can be larger than the ordinary stochastic-encoder code lower bound. Random code lower and upper bounds of the secrecy capacity of the AVWC with constraints are further provided. Based on these results, we further consider a special case of the model, namely severely less noisy AVWC, and give the stochastic-encoder code and random code capacities. It is proved that the stochastic-encoder code capacity of the AVWC with constraints is either equal to or strictly smaller than the corresponding random code capacity, which is consistent with the property of the ordinary AVC. Finally, some numerical examples are presented to better illustrate our capacity results. Compared to the soft covering lemma that requires the codewords to be generated i.i.d., our method has more relaxed requirements regarding codebooks. It is proved that the good codebooks for secure transmission can be generated by choosing codewords randomly from a given type set, which is critical when considering the AVWC with constraints.**

***Index Terms*— Arbitrarily varying wiretap channel, stochastic-encoder code, state constraint, strong secrecy.**

I. Introduction

THE arbitrarily varying channel (AVC) is one of the most challenging communication models. One difficult problem regarding the capacity of the AVC is that it will be affected by different coding schemes and different criteria of decoding error probabilities. Blackwell *et al.* investigated the capacity of the AVC given common randomness [1]. To remove the common randomness, Ahlswede proposed the Elimination Technique in his celebrated paper [2] and provided the average error capacities of the AVC with different coding schemes (deterministic code, stochastic-encoder code, random code) and maximal error capacity of the random code. However, [2] did not fully characterize the average error deterministic code capacity of the AVC since Ahlswede did not give a sufficient and necessary condition for the average error deterministic code capacity to be positive.

The AVC is a discrete memoryless channel with channel states. The state of the channel is selected randomly and the probability mass function of the channel states is unknown. AVC with non-causal state sequence was investigated in [3]. Recently, the arbitrarily varying broadcast/degraded broadcast channels with causal side information at the encoder were investigated in [4] and [5], respectively. Inner and outer bounds for both random code and deterministic code capacities were established. In [6], the AVC with colored Gaussian noise was studied and the deterministic code and random code capacities were provided. An alternative interpretation of the AVC is a communication model with a jammer who controls the channel states and wants to disrupt the communication. For real communication models, it is reasonable to impose constraints on the sender and the jammer. In [7], [8], Csiszár and Narayan presented the random code capacity and the deterministic code capacity of the AVC with input and state constraints, respectively. For an ordinary AVC free of constraints, the deterministic code capacity is either zero or equal to the corresponding random code capacity, which is called Ahlswede's dichotomy [2]. Csiszár and Narayan established a sufficient and necessary condition for the deterministic code capacity of the AVC to be positive, namely that the AVC is non-symmetrizable. However, as shown in [8], the state constraint affects the behavior of the deterministic code capacity of the AVC. Specifically, the deterministic code capacity of the AVC with constraints can be positive even if the channel is symmetrizable. The AVC with constraints and side information at the encoder was studied in [9] and the bounds on the random code capacity and deterministic code capacity were established.

The arbitrarily varying wiretap channel (AVWC) considers the secrecy transmission problem over the AVC. The wiretap channel was first introduced by Wyner [10] and then extended in [11]. A random code lower bound of the secrecy capacity of the AVWC with respect to weak secrecy was provided in [12]. Reference [12] also provided the random code secrecy capacity of the strongly degraded AVWC with a best eavesdropper. In [13], Bjelaković *et al.* studied a special class of the AVWC, where the channel has a 'best eavesdropper', and a random code lower bound of the capacity of the AVWC with respect to strong secrecy was established. Goldfeld *et al.* considered the AVWC with type constraint, where only the

Manuscript received July 25, 2021; revised December 9, 2021; accepted March 11, 2022. Date of publication March 23, 2022; date of current version June 15, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61871264. An earlier version of this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT) 2021 [DOI: 10.1109/ISIT45174.2021.9517973]. *(Corresponding author: Yuan Luo.)*

Yiqi Chen, Chenhao Ying, and Yuan Luo are with the Department of Computer Science and Engineering and the MoE Key Laboratory of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: chenyiqi@sjtu.edu.cn; yingchh1565@sjtu.edu.cn; yuanluo@sjtu.edu.cn).

Dan He is with the Intel Asia-Pacific Research and Development Ltd., Shanghai 200241, China (e-mail: dan.h.he@intel.com).

Communicated by M. Kobayashi, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2022.3161808

state sequences in a set of type sets will occur [14]. The authors of [14] developed a stronger version of Wyner's soft covering lemma [15] and then provided the random code lower and upper bounds of the AVWC with type constraint states with respect to semantic secrecy. Wiese *et al.* established the multi-letter random code secrecy capacity of the AVWC with respect to strong secrecy [16]. In [17], Janda *et al.* provided the random code secrecy capacity of the strongly degraded AVWC with non-causal side information at the jammer and a best eavesdropper. They further provided the multi-letter random code secrecy capacity for the case that the strong degradedness does not hold. In [18], Nötzel *et al.* first discussed the relationship between the stochastic-encoder code secrecy capacity of the AVWC and the corresponding random code. It is proved that the stochastic-encoder code capacity of the AVWC free of constraints equals to its corresponding random code capacity if the main channel is non-symmetrizable.

This paper considers the strong secrecy capacity of the AVWC with general input constraint and state constraint. Two stochastic-encoder code lower bounds of the secrecy capacity without common randomness are provided. The achievability proof of the first bound is based on the coding scheme used in [8]. The strong secrecy is proved with the help of Csiszár's almost independent coloring lemma [19] since the codebooks are not generated i.i.d.. The achievability proof of the second lower bound applies a superposition stochastic coding scheme, which is also used in [20] but for an arbitrarily varying broadcast channel with degraded message sets. A numerical example shows that the superposition coding scheme achieves a higher secrecy achievable rate in some cases. We further give random code lower and upper bounds of the secrecy capacity of the AVWC with input and state constraints by applying the superposition random coding scheme. Finally, based on these results, a special case of the AVWC model, namely severely less noisy AVWC, is considered and the stochastic-encoder code and random code secrecy capacities with constraints are presented. We find that the stochastic-encoder code secrecy capacity of the AVWC with constraints can be strictly smaller than the corresponding random code secrecy capacity, which is consistent with the property of the ordinary AVC. The highlight of our paper is as follows.

1) Two stochastic-encoder code lower bounds and two random code lower bounds of the secrecy capacity of the AVWC with constraints using different realization methods are provided.
2) We introduce superposition coding into both stochastic-encoder coding scheme and random coding scheme. A numerical example is used to discuss the gain of using superposition coding.
3) We prove that the stochastic-encoder code secrecy capacity of the AVWC with constraints can be strictly smaller than the corresponding random code secrecy capacity, which is consistent with the property of the ordinary AVC with constraints.

Based on the above paragraph, the remainder of this paper is organized as follows. Section II introduces the AVWC model considered in this paper and some basic notations and definitions. In Section III, Proposition 1 and Theorem 5 propose the stochastic-encoder code lower bound and superposition stochastic-encoder code lower bound of the secrecy capacity of the AVWC with constraints, respectively; Theorem 6 and Proposition 3 provide the random code lower and upper bounds of the secrecy capacity and Theorem 7 provides the capacity results of the severely less noisy AVWC with constraints. The proof of Proposition 1 and Theorem 5 are provided in Sections IV and V, respectively. Proposition 3 and Theorems 6, 7 are proved in the appendix. In Section VI, three numerical examples are provided to better illustrate our capacity results. Section VII concludes this paper.

The common part of this paper and our conference paper [21] only includes part of the result in Proposition 1.

II. Models

Throughout this paper, random variables and sample values are denoted by capital letters and lowercase letters, respectively. Alphabets and sets are denoted by calligraphic letters. $\mathcal{P}(\mathcal{X})$ denotes the set of probability distributions on a finite alphabet $\mathcal{X}$. Capital and lowercase letters in boldface represent random sequences and sample sequences with length n, e.g. $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

In [8], the deterministic code capacities of the AVC with or without state constraint were derived by using codewords selected from a fixed *type* [19]. The type P_X of a sequence $\boldsymbol{x}$ is defined as

$$P_X(x) = \frac{N(x|\boldsymbol{x})}{n} \qquad \text{for all } x \in \mathcal{X}, \tag{1}$$

where $N(x|\boldsymbol{x})$ is the number of the occurrences of symbol x in the sequence $\boldsymbol{x}$. The set of sequences of type P_X in $\mathcal{X}^n$ is denoted by $T^n_{P_X}$. More definitions and properties are provided in [19].

A. Channel Description

Definition 1 (Definition of the AVWC): Let $\mathcal{X}$ be the finite input alphabet, $\mathcal{S}$ be the finite state alphabet and $\mathcal{Y}, \mathcal{Z}$ be finite output alphabets. An AVWC is a set of channels $(\mathcal{W}, \mathcal{E})$ where $\mathcal{W} = \{W_s : s \in \mathcal{S}\}$ and $\mathcal{E} = \{E_s : s \in \mathcal{S}\}$ satisfying

$$W^n_{\boldsymbol{s}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W_{s_i}(y_i|x_i) = \prod_{i=1}^n W(y_i|x_i, s_i),$$
$$E^n_{\boldsymbol{s}}(\boldsymbol{z}|\boldsymbol{x}) = \prod_{i=1}^n E_{s_i}(z_i|x_i) = \prod_{i=1}^n E(z_i|x_i, s_i).$$

Here W_s is the main channel and E_s is the wiretap channel.

In this paper we only consider the discrete memoryless channel. To characterize the capacity results of the AVWC with constraints, we also need to consider the convex hull $\overline{\mathcal{W}}$ of a given AVC $\mathcal{W}$. For a given state distribution $q \in \mathcal{P}(\mathcal{S})$, the channel W_q is defined as $W_q(y|x) \triangleq \sum_{s\in\mathcal{S}} q(s)W(y|x,s)$ for any $x \in \mathcal{X}, y \in \mathcal{Y}$, and then the convex hull of the given AVC $\mathcal{W}$ is defined as $\overline{\mathcal{W}} = \{W_q : q \in \mathcal{P}(\mathcal{S})\}$. One can regard $\overline{\mathcal{W}}$ as an AVC with state space $\mathcal{P}(\mathcal{S})$. It follows that $\mathcal{W} \subset \overline{\mathcal{W}}$ since q can be a one-point distribution, i.e. $q(s) = 1$ for $s \in \mathcal{S}$ and $q(s') = 0$ for any $s' \neq s$. Let Y_q be the output of the channel W_q and Y_s be the output of the channel W_s.

One interpretation of the AVC model is a communication system in the presence of a jammer who controls the channel

state and wants to disrupt the communication. For a real communication system, it is reasonable to use functions ψ and l to impose constraints on the sequence of channel input and sequence of channel states, respectively.

Definition 2 (Constraints on the AVWC): Let $\psi : \mathcal{X} \to [0, +\infty)$ and $l : \mathcal{S} \to [0, +\infty)$ be given constraint functions on the input alphabet $\mathcal{X}$ and the state alphabet $\mathcal{S}$, respectively. For given input sequence $\boldsymbol{x}$ and state sequence $\boldsymbol{s}$, define

$$\psi^n(\boldsymbol{x}) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \psi(x_i), l^n(\boldsymbol{s}) = \frac{1}{n}\sum\nolimits_{i=1}^{n} l(s_i). \quad (2)$$

Furthermore, we assume $\min_{x\in\mathcal{X}} \psi(x) = \min_{s\in\mathcal{S}} l(s) = 0$. Let $\mathcal{S}^n(\Lambda)$ be the set of state sequences satisfying $l^n(\boldsymbol{s}) \leq \Lambda$ for some $\Lambda > 0$. An AVWC with state constraint Λ is an AVWC that only $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ will occur. For the input constraint, given message set $\mathcal{M}$ and encoder $f : \mathcal{M} \to \mathcal{X}^n$, the encoder should satisfy $\psi^n(f(m)) \leq \Upsilon$ for some $\Upsilon > 0$ and any $m \in \mathcal{M}$.

Remark 1: Note that for given sequence $\boldsymbol{x}$ with type P_X, the input cost function can be written as $\psi^n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i) = \sum_{x\in\mathcal{X}} P_X(x)\psi(x)$. For a set of input sequences with same type P_X, define $\psi(P_X) \triangleq \sum_{x\in\mathcal{X}} P_X(x)\psi(x)$.

The following definitions describe the stochastic and random coding schemes and the corresponding error criteria considered in this paper.

Definition 3 (Stochastic-Encoder Code (SC) Over the AVWC): A stochastic-encoder code over the AVWC is a tuple (f, ϕ) with a message set $\mathcal{M} = [1 : 2^{nR}]$. The tuple (f, ϕ) consists of a stochastic encoder $f : \mathcal{M} \to \mathcal{X}^n$ such that for each message $m \in \mathcal{M}$, $\boldsymbol{X}(m)$ is a sequence distributed according to $f(\cdot|m)$ and a deterministic decoder $\phi : \mathcal{Y}^n \to \mathcal{M}$. The average error probability under state sequence $\boldsymbol{s}$ of the code (f, ϕ) is defined as

$$\bar{\lambda}^{SC}(\mathcal{W}, f, \phi, \boldsymbol{s}) = \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \lambda_m^{SC}(\mathcal{W}, f, \phi, \boldsymbol{s}),$$

where

$$\lambda_m^{SC}(\mathcal{W}, f, \phi, \boldsymbol{s}) = \sum\nolimits_{\boldsymbol{x}\in\mathcal{X}^n} f(\boldsymbol{x}|m) \cdot (1 - W^n(\phi^{-1}(m)|\boldsymbol{x}, \boldsymbol{s})).$$

Definition 4 (Random Code (RC) Over the AVWC): A random code over the AVWC is a triple (F, Φ, Γ) with message set $\mathcal{M}$. Here (F, Φ) is a random encoder/decoder pair distributed on a set of stochastic-encoder codes $(f^\gamma, \phi^\gamma)_{\gamma\in\mathcal{I}}$ with index set $\mathcal{I}$ and distribution μ; Γ is the common randomness random variable on $\mathcal{I}$ satisfying $Pr\{\Gamma = \gamma\} = \mu(\gamma)$ for $\gamma \in \mathcal{I}$. The average error probability under state sequence $\boldsymbol{s}$ of the code (F, Φ, Γ) is defined as

$$\bar{\lambda}^{RC}(\mathcal{W}, F, \Phi, \boldsymbol{s}) = \sum_{\gamma\in\mathcal{I}} \mu(\gamma)\bar{\lambda}^{SC}(\mathcal{W}, f^\gamma, \phi^\gamma, \boldsymbol{s}).$$

The difference between the stochastic-encoder code and random code is the existence of the common randomness in random code. For the stochastic-encoder code, there is only local randomness at the sender side. For the random code, there is an additional common randomness between the sender and the receiver. Each time, to send a message, the common randomness uniformly selects a stochastic-encoder code at random and reveals it to the sender and the receiver. Using the coding schemes defined above, we move on to the definitions of the secrecy achievable rate and the strong secrecy capacity of the AVWC with constraints.

Definition 5 (Achievable Rate and Capacity of the AVWC): A non-negative real number R is said to be secrecy achievable rate by stochastic-encoder code with input constraint Υ and state constraint Λ if for any $\tau, \varepsilon > 0$ and sufficiently large n, there exists a stochastic-encoder code (f, ϕ) such that

$$\begin{aligned} &\frac{1}{n}\log|\mathcal{M}| \geq R - \tau, \\ &\max_{\boldsymbol{s}\in\mathcal{S}^n(\Lambda)} \bar{\lambda}^{SC}(\mathcal{W}, f, \phi, \boldsymbol{s}) \leq \varepsilon, \\ &\max_{\boldsymbol{s}\in\mathcal{S}^n(\Lambda)} I(M; \boldsymbol{Z}_{\boldsymbol{s}}) < \varepsilon, \end{aligned}$$

where $I(M; \boldsymbol{Z}_{\boldsymbol{s}})$ is the information leakage under state sequence $\boldsymbol{s}$. The stochastic-encoder code secrecy capacity of the AVWC with constraints is the supremum of the stochastic-encoder code secrecy achievable rate. Similarly, a real number R is said to be secrecy achievable rate by random code with input constraint Υ and state constraint Λ if for any $\tau, \varepsilon > 0$ and sufficiently large n, there exists a random code (F, Φ, Γ) such that

$$\begin{aligned} &\frac{1}{n}\log|\mathcal{M}| \geq R - \tau, \\ &\max_{\boldsymbol{s}\in\mathcal{S}^n(\Lambda)} \bar{\lambda}^{RC}(\mathcal{W}, F, \Phi, \boldsymbol{s}) \leq \varepsilon, \\ &\max_{\boldsymbol{s}\in\mathcal{S}^n(\Lambda)} I(M; \boldsymbol{Z}_{\boldsymbol{s}}|\Gamma) < \varepsilon, \end{aligned}$$

where $I(M; \boldsymbol{Z}_{\boldsymbol{s}}|\Gamma)$ is the information leakage under state sequence $\boldsymbol{s}$ with respect to random code. The random code secrecy capacity of the AVWC with constraints is the supremum of the random code secrecy achievable rate.

B. Symmetrizability

For an ordinary AVC, a celebrated result called Ahlswede's Dichotomy, shown by Ahlswede in his paper [2] asserts that the deterministic code capacity and stochastic-encoder code capacity of an ordinary AVC are either equal to its random code capacity or else are zero. A necessary and sufficient condition for the positivity of the capacity of the AVC is presented in [8], [22]. The following definition describes the 'symmetrizable AVC', which leads to the zero capacity of an AVC free of constraints.

Definition 6 (Symmetrizable AVC): An AVC is symmetrizable-$\mathcal{X}$ if for some channel $T : \mathcal{X} \to \mathcal{S}$,

$$\sum_{s\in\mathcal{S}} W(y|x, s)T(s|x') = \sum_{s\in\mathcal{S}} W(y|x', s)T(s|x) \quad \text{for every } x, x', y. \quad (3)$$

The symmetrizable-$\mathcal{X}$ AVC is described as a poor channel [8], [9] since the decoder couldn't distinguish between two possible codewords. See the following example in [8].

Example 1: Let $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$. The arbitrarily varying channel is specified by

$$Y = X + S \mod 2.$$

It can be proved that the channel satisfies the symmetrizable-$\mathcal{X}$ condition in Definition 6 if and only if $T : \mathcal{X} \to \mathcal{S}$ is a symmetric channel, i.e.

$$T = \begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix} \qquad \text{for } 0 \le e \le 1.$$

For such a channel, the jammer selects a state sequence by $T^n(\boldsymbol{s}|\boldsymbol{x}_2) = \prod_{i=1}^n T(s_i|x_{2i})$ with codeword $\boldsymbol{x}_2$ when codeword $\boldsymbol{x}_1$ is used by the sender. Due to the property of symmetrizable-$\mathcal{X}$, the average channel satisfies

$$\begin{aligned} &W^n_{T(\cdot|\boldsymbol{x}_2)}(\boldsymbol{y}|\boldsymbol{x}_1) \\ &= \prod_{i=1}^n W_{T(\cdot|x_{2i})}(y_i|x_{1i}) \\ &= \prod_{i=1}^n \sum_s W(y_i|x_{1i}, s)T(s|x_{2i}) \\ &= \prod_{i=1}^n \sum_s W(y_i|x_{2i}, s)T(s|x_{1i}) = W^n_{T(\cdot|\boldsymbol{x}_1)}(\boldsymbol{y}|\boldsymbol{x}_2) \end{aligned}$$

for any $\boldsymbol{y} \in \mathcal{Y}^n$. In this case, the receiver cannot distinguish between codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ from the average channel. For a deterministic decoder ϕ, let disjoint sets $\mathcal{D}_1$ and $\mathcal{D}_2$ be the decoding sets of codewords $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively. Suppose $W^n_{T(\cdot|\boldsymbol{x}_2)}(\mathcal{D}_1|\boldsymbol{x}_1) = 1-\epsilon$ for some $0 < \epsilon < 1/2$. By the above equalities, we have $W^n_{T(\cdot|\boldsymbol{x}_2)}(\mathcal{D}_1|\boldsymbol{x}_1) = W^n_{T(\cdot|\boldsymbol{x}_1)}(\mathcal{D}_1|\boldsymbol{x}_2) = 1-\epsilon$ and hence $W^n_{T(\cdot|\boldsymbol{x}_1)}(\mathcal{D}_2|\boldsymbol{x}_2) < \epsilon < 1/2$.

Note that the random code capacity of the example is also 0. Here, we only use this example to demonstrate the ambiguousness caused by the symmetrizability, which also holds for other symmetrizable AVCs. Coding schemes without common randomness cannot achieve reliable communication over an AVWC free of constraints with a symmetrizable main channel, and hence the corresponding stochastic-encoder code capacity is 0. However, as mentioned in [8], the capacity of the AVC with state constraint Λ may be positive even if it is symmetrizable. For a given AVC with state constraint and input distribution P_X, the positivity of the channel capacity depends on whether Λ is larger or smaller than

$$\Lambda_0(P_X) = \min_{T \in \mathcal{T}} \sum_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} P_X(x)T(s|x)l(s), \tag{4}$$

where $\mathcal{T}$ is the set of channels $T : \mathcal{X} \to \mathcal{S}$ satisfying formula (3). Furthermore, we follow the setting in [8] that $\Lambda_0(P_X) = \infty$ if $\mathcal{T} = \emptyset$. In fact $\Lambda_0(P_X)$ is the minimal cost, given input distribution P_X, for the jammer to symmetrize an AVC. Thus, if the minimal cost is less than the state constraint Λ, the jammer can always choose some $T : \mathcal{X} \to \mathcal{S}$ that symmetrizes the channel, leading to zero capacity.

In [11], the prefixed channel argument suggests that a virtual pre-channel can improve the secrecy capacity of the wiretap channel in some cases. Accordingly, the definition of the symmetrizable AVC is extended as follows.

Definition 7 (Symmetrizable AVC in the Presence of the Prefixed Channel): An AVC with prefixed channel $P_{X|U}$ is symmetrizable-$\mathcal{U}$ if for some channels $T : \mathcal{X} \to \mathcal{S}$ and $T'(s|u) = \sum_{x \in \mathcal{X}} P_{X|U}(x|u)T(s|x)$,

$$\sum_{s \in \mathcal{S}} W(y|u, s)T'(s|u') = \sum_{s \in \mathcal{S}} W(y|u', s)T'(s|u) \quad \text{for every } u, u' \in \mathcal{U}, y \in \mathcal{Y}, \tag{5}$$

where $W(y|u, s) = \sum_{x \in \mathcal{X}} W(y|x, s)P_{X|U}(x|u)$.

The set of channels T satisfying Definition 7 is different from the set of channels T satisfying Definition 6 once the prefixed channel $P_{X|U}$ is specified, and we will discuss their difference later in Remark 2. For simplicity, denote the set of channels satisfying the symmetrizable-$\mathcal{U}$ condition by $\mathcal{T}'$. The corresponding minimal cost for the jammer is defined by

$$\Lambda_1(P_{UX}) = \min_{T' \in \mathcal{T}'} \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}} P_U(u)T'(s|u)l(s), \tag{6}$$

where $\mathcal{T}'$ is the set of channels T' satisfying formula (5). Similarly, $\Lambda_1(P_{UX}) = \infty$ if $\mathcal{T}' = \emptyset$. For the AVWC with input constraint, it follows that

$$\begin{aligned} \psi^n(\boldsymbol{x}) &= \frac{1}{n}\sum_{i=1}^n \psi(x_i) = \sum_{x \in \mathcal{X}} \psi(x)P_X(x) \\ &= \sum_{x \in \mathcal{X}} \psi(x) \sum_{u \in \mathcal{U}} P_{UX}(u, x) \triangleq \psi(P_{UX}). \end{aligned} \tag{7}$$

Similar to the original channel [8, Lemma 1], the error probability of the AVC with a prefixed channel is nontrivial if $\Lambda_1(P_{UX}) < \Lambda$. This can be easily extended by Definition 7 and the proof of Lemma 1 in [8]. Let $\Lambda_1^*(U, X) = \sup_{P_{UX}} \Lambda_1(P_{UX})$. If $\Lambda_1^*(U, X) > \Lambda$, there always exists at least one joint type P_{UX} satisfying $\Lambda_1(P_{UX}) > \Lambda$, then we can construct a code that is achievable over the AVWC with state constraint.

Remark 2 (Relation Between $\Lambda_1(P_{UX})$ and $\Lambda_0(P_X)$): In this remark we talk about the relation between

$$\Lambda_0(P_X) = \min_{T \in \mathcal{T}} \sum_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} P_X(x)T(s|x)l(s) \tag{8}$$

and

$$\Lambda_1(P_{UX}) = \min_{T' \in \mathcal{T}'} \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}} P_U(u)T'(s|u)l(s). \tag{9}$$

Note that the prefixed channel $P_{X|U}$ is selected by the sender and fixed during the transmission. Hence, the channel $T'(s|u) = \sum_{x \in \mathcal{X}} P_{X|U}(x|u)T(s|x)$. Then formula (9) can be written as

$$\Lambda_1(P_{UX}) = \min_{T \in \mathcal{T}_1} \sum_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} P_U(u)P_{X|U}(x|u)T(s|x)l(s), \tag{10}$$

where $\mathcal{T}_1$ is the set of channels T satisfying the symmetrizable-$\mathcal{U}$ condition. Furthermore, for convenience, we write the set of channels $\mathcal{T}$ in formula (8) by $\mathcal{T}_0$. In [18], the authors considered the stochastic-encoder code capacity of a non-symmetrizable AVWC. One interesting result, as stated in Example 1 in [18], is that the non-symmetrizable AVWC may be symmetrizable after adding a prefixed channel. Hence, there may exist a channel T such that $T \notin \mathcal{T}_0$ but $T \in \mathcal{T}_1$. On the other hand, it can be proved that an AVWC

is always symmetrizable-$\mathcal{U}$ if it is symmetrizable-$\mathcal{X}$. As a result, we have $\mathcal{T}_0 \subseteq \mathcal{T}_1$. For given distributions P_{UX} and P_X such that $P_X(x) = \sum_u P_{UX}(u,x)$, it follows that $\Lambda_1(P_{UX}) \leq \Lambda_0(P_X)$. In the case of the AVWC with state constraint, the capacity can still be positive even if the AVWC is symmetrizable after adding a prefixed channel as long as $\Lambda_1^*(U,X) > \Lambda$ is satisfied. Note that the input distribution in the presence of a prefixed channel is limited to $\Lambda_1(P_{UX}) \geq \Lambda$ when the non-symmetrizable AVWC is symmetrizable after adding the prefixed channel. However, by setting $U = X$, one can guarantee that the prefixed channel will never reduce the achievable rate.

Next we consider the symmetrizability of the AVWC regarding the superposition coding scheme. Superposition coding [23] is a layered coding approach that for joint distribution $P_{VX} = P_V P_{X|V}$, there exists a 'cloud center' codebook $\mathcal{C}_C = \{\boldsymbol{v}(i)\}_{i=1}^{N_C}$ generated by distribution P_V for some $N_C > 0$, and the codewords in $\mathcal{C}(i) = \{\boldsymbol{x}(i,j)\}_{j=1}^{N}, N > 0$, are generated by distribution $P_{X|V}$ for each 'cloud center' $\boldsymbol{v}(i) \in \mathcal{C}_C$. The AVWC can be interpreted as an arbitrarily varying broadcast channel (AVBC) if we regard the eavesdropper as another receiver. Hence, the argument about the symmetrizability of the AVBC is also applicable to the AVWC. Accordingly, the definitions of symmetrizable channels and the minimum average cost of the jammer are extended as follows.

Definition 8 (Definition 9 in [20]): An AVC $\mathcal{W}$ is symmetrizable-$\mathcal{X}|\mathcal{V}$ if there exists some channels $\widetilde{T} : \mathcal{V} \times \mathcal{X} \to \mathcal{S}$ satisfying

$$\sum_{s\in\mathcal{S}} W(y|x,s)\widetilde{T}(s|v,x') = \sum_{s\in\mathcal{S}} W(y|x',s)\widetilde{T}(s|v,x) \quad (11)$$

for all v, x, x', s, y with joint distribution P_{VX} satisfying $P_{VX}(v,x) = P_V(v)P_{X|V}(x|v)$ and $\min_{v,x} P_{VX}(v,x) > 0$. The set of channels $\widetilde{T}$ that $\mathcal{X}|\mathcal{V}$-symmetrizes $\mathcal{W}$ is denoted by $\widetilde{\mathcal{T}}$.

The corresponding minimal average cost of the jammer for given distribution P_{VX} is

$$\widetilde{\Lambda}_0(P_{VX}) \triangleq \min_{\widetilde{T}\in\widetilde{\mathcal{T}}} \sum_{v\in\mathcal{V}}\sum_{x\in\mathcal{X}}\sum_{s\in\mathcal{S}} P_{VX}(v,x)\widetilde{T}(s|v,x)l(s). \quad (12)$$

Similarly, define

$$\widetilde{\Lambda}_0^*(V,X) \triangleq \max_{P_{VX}} \widetilde{\Lambda}_0(P_{VX}).$$

When considering the arbitrarily varying wiretap channel, define the minimal average cost of the jammer in the presence of a prefixed channel as

$$\widetilde{\Lambda}_1(P_{VUX}) \triangleq \min_{\widetilde{T}\in\widetilde{\mathcal{T}}} \sum_{v\in\mathcal{V}}\sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} P_{VU}(v,u)\widetilde{T}(s|v,x)P_{X|U}(x|u)l(s),$$

where $P_{VUX}(v,u,x) = P_{VU}(v,u)P_{X|U}(x|u)$, and $\widetilde{\Lambda}_1(P_{VUX}) = \infty$ if $\widetilde{\mathcal{T}} = \emptyset$. We further define

$$\widetilde{\Lambda}_1^*(V,U,X) \triangleq \max_{P_{VUX}} \widetilde{\Lambda}_1(P_{VUX}).$$

The definitions of the communication model and the coding schemes used in this paper have now been introduced. In the next section, we present some existing work about AVWC without constraints.

C. Related Work

In the following paragraphs, we review some existing results on the arbitrarily varying wiretap channel.

1) Random Code Results: [12] defined a special case of the AVWC such that

- The main channel and the wiretap channel have separate state spaces $\mathcal{S}_y$ and $\mathcal{S}_z$ corresponding to the main channel and the wiretap channel, respectively;
- The eavesdropper has a best channel E_{q^*} such that $X \to Z_{q^*} \to Z_{s_z}$ for any $s_z \in \mathcal{S}_z$;
- The AVWC is strongly degraded such that $X \to Y_q \to Z_{q'}$ holds for any $q \in \mathcal{P}(\mathcal{S}_y)$ and $q' \in \mathcal{P}(\mathcal{S}_z)$.

Based on the above conditions, [12] established the single-letter random code secrecy capacity of the AVWC with respect to weak secrecy, i.e. the information leakage $\max_{\boldsymbol{s}\in\mathcal{S}^n} \frac{1}{n} I(M; \boldsymbol{Z_s}|\Gamma) < \varepsilon$ for every $\varepsilon > 0$ and sufficiently large n.

Theorem 1 ([12]): The random code secrecy capacity of the strongly degraded AVWC with independent states is

$$C^{RC} = \max_{P_X\in\mathcal{P}(\mathcal{S})} \left[\min_{q\in\mathcal{P}(\mathcal{S}_y)} I(X;Y_q) - \max_{q'\in\mathcal{P}(\mathcal{S}_z)} I(X;Z_{q'})\right],$$

provided the AVWC has a best channel for the eavesdropper.

In [13], the authors considered a special case of the AVWC assuming the existence of a best channel to the eavesdropper and established the following lower bound of the random code secrecy capacity with respect to strong secrecy.

Theorem 2 ([13]): For the AVWC with a best eavesdropper, the random code (RC) secrecy capacity is lower bounded by

$$C^{RC} \geq \max_{P_X\in\mathcal{P}(\mathcal{X})} \left[\min_{q\in\mathcal{P}(\mathcal{S})} I(X;Y_q) - \max_{q\in\mathcal{P}(\mathcal{S})} I(X;Z_q)\right]$$

with joint distributions $P_X(x)q(s)W(y|x,s)$ and $P_X(x)q(s)E(z|x,s)$ for $x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}, s \in \mathcal{S}$.

The above results both have strong assumptions on the structure of the AVWC. In [14], the authors considered a general AVWC model with state types constraints without any assumptions on the structure of the channel and established the following random code lower bound of the secrecy capacity with respect to semantic secrecy.

Theorem 3 ([14]): For any convex and closed $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{S})$, the random code semantic secrecy capacity of the AVWC with $\mathcal{Q}$-constraint is lower bounded as

$$C^{RC} \geq \max_{P_{UX}} \left[\min_{q^1\in\mathcal{Q}} I(U;Y_q) - \max_{q^2\in\mathcal{Q}} I(U;Z|S)\right]$$

where the mutual information terms are calculated with respect to joint distributions $P_{UX} \cdot q^j \cdot W_{Y|XS} \cdot E_{Z|XS}$ for $j = 1, 2$, and $|\mathcal{U}| \leq |\mathcal{X}|$.

The result in Theorem 3 was established with respect to semantic secrecy, which doesn't make any assumptions about the message distribution. Therefore, it is severer than strong secrecy requirement. The authors of [14] further provided a random code upper bound of the general AVWC with $\mathcal{Q}$-constraint. Comparing to the results in [14], our works focus on both stochastic-encoder and random code bounds of the general AVWC with additional input and state constraints imposed on the sender and jammer, respectively.

2) Stochastic-Encoder Code Results: In [18], the authors discussed the relationship between the random code secrecy capacity and the stochastic-encoder code secrecy capacity without any constraints. They characterized the stochastic-encoder code capacity of the AVWC without constraints as follows.

Theorem 4 ([18]): If the random code (RC) secrecy capacity of the AVWC $C^{RC} > 0$, the stochastic code secrecy capacity is given by

$$C^{SC} = C^{RC}$$

if and only if the main channel is non-symmetrizable-$\mathcal{X}$. If the main channel is symmetrizable-$\mathcal{X}$, $C^{SC} = 0$.

The result in 4 is consistent with the AVC dichotomy shown by Ahlswede in [2]. In this paper, we further investigate the capacity of symmetrizable AVWC with constraints and prove that the stochastic-encoder secrecy capacity of the AVWC with constraints can be positive but strictly smaller than the random code secrecy capacity.

3) Summary: Although a multi-letter characterization of the random code secrecy capacity of the AVWC with respect to strong secrecy was established in [16], the single-letter form of the secrecy capacity is still an open problem. In this paper, we focus on the strong secrecy communication problem over the general AVWC with input and state constraints without any ohter assumptions.

III. Main Results

This section provides the main capacity results of this paper. The first stochastic achievable rate result is given in Proposition 1 based on the coding scheme in [8]. In Theorem 5, an achievable rate established by superposition stochastic coding scheme is provided. We further discuss the similarities and differences between the superposition stochastic coding scheme and random coding scheme. A Lower bound of the random code secrecy capacity of the AVWC with constraints is established in Theorem 6 by applying superposition random coding scheme, and the upper bound is provided in Proposition 3. Finally, we consider a special case of the AVWC, namely severely less noisy AVWC with constraints, and provide the corresponding stochastic-encoder code and random code capacities in Theorem 7.

For notation convenience, we use the subscript q to represent the distribution of the random variable S contained in the mutual information term. For example, for random variables X, S, Y such that $Pr\{X = x, S = s, Y = y\} = P_X(x) \cdot q(s) \cdot W(y|x, s)$, the mutual information of X and Y conditioning on S is written as $I_q(X;Y|S) = \sum_{s\in\mathcal{S}} q(s) I(X;Y|S = s)$. Now the first main result of this paper can be given.

A. Stochastic-Encoder Code Bounds

Proposition 1 (Lower Bound of the Stochastic-Encoder Code Secrecy Capacity of the AVWC With Constraints): If $\Lambda_1^*(U, X) > \Lambda$, the stochastic-encoder code (SC) secrecy capacity C^{SC} of the AVWC with state constraint Λ and input constraint Υ is lower bounded by

$$C^{SC} \geq \max_{P_{UX}\in\mathcal{P}_{\Upsilon,\Lambda}(\mathcal{U},\mathcal{X})} [\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q) - \max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(U;Z|S)] \tag{13}$$

under joint distribution $P_{UX} \times W_{Y_q|X}$ and $P_{UX} \times q' \times E_{Z|XS}$, where the state distribution set $\mathcal{P}_\Lambda(\mathcal{S}) = \{q \in \mathcal{P}(\mathcal{S}) : \mathbb{E}[l(S)] = \sum_{s\in\mathcal{S}} q(s)l(s) \leq \Lambda\}$, the input distribution set $\mathcal{P}_{\Upsilon,\Lambda}(\mathcal{U},\mathcal{X}) = \{P_{UX} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}) : \psi(P_{UX}) \leq \Upsilon, \Lambda_1(P_{UX}) \geq \Lambda\}$, and $I_{q'}(U;Z|S) = \sum_s q'(s) I(U;Z_s)$, $\psi(P_{UX}) = \sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} P_U(u)P_{X|U}(x|u)\psi(x)$.

The proof is provided in Section IV and also our conference paper [21]. Similar to the original AVWC, we first set $U = X$ in formula (13) and prove the achievability. Then Proposition 1 is proved by applying the standard prefixed channel argument. When $\max_s l(s) \leq \Lambda, \max_x \psi(x) \leq \Upsilon$, the constraints are inactive and the result reduces to the lower bound of the capacity of the non-symmetrizable AVWC free of constraints considering the definition that $\Lambda_1(P_{UX}) = \infty$.

In [14], the authors considered the AVWC with type constraint states. Note that Proposition 1 in this paper and the result in [14] do not cover each other since the achievable rate result in [14] is with respect to semantic secrecy with common randomness.

We further give a proposition about a lower bound of the random code secrecy capacity of the AVWC with input and state constraints.

Proposition 2 (Lower Bound of the Random Code Secrecy Capacity of the AVWC With Constraints): The random code (RC) secrecy capacity C^{RC} of the AVWC with input constraint Υ and state constraint Λ is lower bounded by

$$C^{RC} \geq \max_{\substack{P_{UX}:\\ \psi(P_{UX})\leq\Upsilon}} [\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q) - \max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(U;Z|S)]$$

under joint distribution $P_{UX} \times W_{Y_q|X}$ and $P_{UX} \times q' \times E_{Z|XS}$, where $\mathcal{P}_\Lambda(\mathcal{S}) = \{q \in \mathcal{P}(\mathcal{S}) : \mathbb{E}[l(S)] = \sum_{s\in\mathcal{S}} q(s)l(s) \leq \Lambda\}, \psi(P_{UX}) = \sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} P_U(u)P_{X|U}(x|u)\psi(x)$.

The proof is provided in Appendix A. The random code capacity of the arbitrarily varying multiple access channel with input and state constraints was investigated in [24]. The result in Proposition 2 follows directly when applying the technique used in the forward proof in [24] and Lemma 5 in Section IV. The main difference in the lower bounds between Proposition 1 and Proposition 2 is the range of input distribution. There is an additional constraint on the input distribution in Proposition 1, i.e. $\Lambda_1(P_{UX}) \geq \Lambda$. Hence, if the input distribution that maximizes the random code lower bound in Proposition 2 violates the condition that $\Lambda_1(P_{UX}) \geq \Lambda$, the stochastic-encoder code lower bound of the AVWC with constraints can be strictly smaller than the corresponding random code lower bound.

B. Superposition Stochastic-Encoder Code Bounds

In [25], superposition coding scheme was applied to the wiretap channel with input constraint. It is proved that an additional auxiliary random variable is necessary to achieve the secrecy capacity of the wiretap channel in some cases. Here

we apply the superposition coding scheme to the AVWC with constraints and derive a lower bound of the secrecy capacity with respect to strong secrecy.

Theorem 5: For an AVWC with input constraint Υ and state constraint Λ, if $\widetilde{\Lambda}_1^*(V,U,X) > \Lambda$, the stochastic-encoder code secrecy capacity C^{SC} is lower bounded by

$$C^{SC} \geq \max_{P_{VUX} \in P_{\Upsilon,\Lambda}(\mathcal{V},\mathcal{U},\mathcal{X})} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q|V) - \max_{q' \in \mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(U;Z|S,V)],$$

where $P_{\Upsilon,\Lambda}(\mathcal{V},\mathcal{U},\mathcal{X}) = \{P_{VUX} \in \mathcal{P}(\mathcal{V}\times\mathcal{U}\times\mathcal{X}) : \psi(P_{VUX}) \leq \Upsilon, \widetilde{\Lambda}_1(P_{VUX}) \geq \Lambda\}, \psi(P_{VUX}) = \sum_{v\in\mathcal{V},u\in\mathcal{U},x\in\mathcal{X}} P_V(v)\ P_{U|V}(u|v)P_{X|U}(x|u)\psi(x)$, and $\mathcal{P}_\Lambda(\mathcal{S}) = \{q \in \mathcal{P}(\mathcal{S}) : \mathbb{E}[l(S)] = \sum_{s\in\mathcal{S}} q(s)l(s) \leq \Lambda\}$.

The proof is given in Section V. The achievability of Theorem 5 is similar to that in [20] for the arbitrarily varying broadcast channel (AVBC) with degraded message set. The main difference is that there is no requirement on the correctness of the second receiver (the eavesdropper in this paper). The 'cloud center' codewords are i.i.d. and uniformly selected from a fixed type set. For each 'cloud center' codeword, the codewords for the private message are then selected from the corresponding conditional type set. Similar to the coding scheme discussed previously, the superposition stochastic-encoder code capacity of the AVWC with constraints can be positive even if the main channel is symmetrizable.

The advantage of applying superposition coding to the wiretap channel was investigated in [25]. In [25], it was proved that the superposition coding scheme can improve the capacity of ordinary wiretap channels in the presence of the input constraint. The argument also holds for the arbitrarily varying wiretap channel with input constraint and state constraint. In Section VI, we give an example that is an arbitrarily varying channel version of the example investigated in [25], which shows that the superposition coding scheme can also improve the achievable rate of the AVWC with input constraint and state constraint.

The superposition coding scheme is somewhat similar to the random coding scheme. In the random coding scheme, the common randomness uniformly selects a code at random from a set of codes before each message is sent. This additional common randomness helps the participants in the communication system in the presence of the jammer. In [2], Ahlswede proposed the Elimination Technique to construct deterministic code from a given random code over the AVC free of constraints with the precondition that the deterministic code capacity of the AVC is positive. Specifically, this random code reduction technique shows that a polynomial number of codes, e.g. n^2, is sufficient for a random code over the AVC to transmit messages reliably. Based on this the sender selects a deterministic code with message set $[1 : n^2]$ as the n_0 length prefixed code with n_0/n vanishing as $n \to \infty$. The new deterministic code is the juxtaposition of the prefixed code and the original random code. However, this technique does not work for the AVC with state constraint. An alternative coding scheme is the superposition coding scheme considered in this paper. As mentioned before, the 'cloud center' of the superposition coding is some meaningless codewords, each corresponding to a stochastic-encoder codebook. The selection of 'cloud center' codewords introduces the additional randomness in the transmission. Instead of encoding the randomness into the prefixed code, the superposition coding scheme encodes the additional randomness and the messages together into codewords with length n. The proof shows that the coding scheme satisfies both reliability and the strong secrecy requirements. The example in Section VI shows that the coding scheme achieves higher achievable rate in some cases.

Theorem 6: The random code secrecy capacity of the AVWC with input constraint Υ and state constraint Λ is lower bounded by

$$C^{RC} \geq \max_{P_{VUX} \in \mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X})} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q|V) - \max_{q' \in \mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(U;Z|V,S)],$$

with distribution $P_{VUX}(v,u,x) = P_V(v)P_{U|V}(u|v)P_{X|U}(x|u)$ for any $V \in \mathcal{V}, u \in \mathcal{U}$ and $x \in \mathcal{X}$, where $\mathcal{P}_\Lambda(\mathcal{S}) = \{q \in \mathcal{P}(\mathcal{S}) : \mathbb{E}[l(S)] = \sum_{s\in\mathcal{S}} q(s)l(s) \leq \Lambda\}$, $\mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X}) = \{P_{VUX} : \psi(P_{VUX}) \leq \Upsilon\}$, $\psi(P_{VUX}) = \sum_{v,u,x} P_V(v)P_{U|V}(u|v)P_{X|U}(x|u)\psi(x)$.

The proof is given in Appendix D-A. The random coding scheme based on joint typicality decoding was applied to the arbitrarily varying channel in [24], [26]. The main difference between the coding schemes in [26] and [24] is that in [26], the letters of each codeword are independent while the codewords in [24] are randomly selected from some fixed type sets due to the input constraint. Here we adopt the random coding scheme in [24], which is also used in the proof of Proposition 2.

Proposition 3: The random code secrecy capacity of the AVWC with input constraint Υ and state constraint Λ is upper bounded by

$$C^{RC} \leq \min_{q,q' \in \mathcal{P}_\Lambda(\mathcal{S})} \max_{P_{VUX} \in \mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X})} [I(U;Y_q|V) - I_{q'}(U;Z|V,S)],$$

with input distribution $P_{VUX}(v,u,x) = P_V(v)P_{U|V}(u|v)$ $P_{X|U}(x|u)$ for any $V \in \mathcal{V}, u \in \mathcal{U}$ and $x \in \mathcal{X}$, where $\mathcal{P}_\Lambda(\mathcal{S}) = \{q \in \mathcal{P}(\mathcal{S}) : \mathbb{E}[l(S)] = \sum_{s\in\mathcal{S}} q(s)l(s) \leq \Lambda\}$, $\mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X}) = \{P_{VUX} : \psi(P_{VUX}) \leq \Upsilon\}$, $\psi(P_{VUX}) = \sum_{v,u,x} P_V(v)P_{U|V}(u|v)P_{X|U}(x|u)\psi(x)$.

The proof is provided in Appendix D-B. As explained in [14], the auxiliary random variable V allows the sender to choose a random mixture of the input distribution since there may not exist a state that is bad for the whole mixture. In the next subsection, we prove that the superposition coding is not necessary in some special cases.

C. Special Case: Severely Less Noisy AVWC With Constraints

For ordinary discrete memoryless channels, a DMC $W : \mathcal{X} \to \mathcal{Y}$ is said to be less noisy than another DMC $E : \mathcal{X} \to \mathcal{Z}$ if $I(U;Y) \geq I(U;Z)$ for any input distribution P_{UX}, where

the joint distribution of (U, X, Y, Z) satisfies $Pr\{U = u, X = x, Y = y, Z = z\} = Pr\{U = u, X = x\}W(y|x)E(z|x)$.

For the arbitrarily varying channel, an AVC $\mathcal{W} : \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$ is said to be *severely less noisy* [27] than $\mathcal{E} : \mathcal{X} \times \mathcal{S} \to \mathcal{Z}$ if $I(U; Y_q) \geq I(U; Z_s)$ for any $q \in \mathcal{P}(\mathcal{S})$ and any $s \in \mathcal{S}$. An AVWC is severely less noisy if the main channel is severely less noisy than the wiretap channel. Then we have the following capacity results.

Theorem 7: The stochastic-encoder code secrecy capacity of a severely less noisy AVWC with input constraint Υ and state constraint Λ is

$$C^{SC} = \max_{\substack{P_{VX} \in \\ \mathcal{P}_{\Upsilon,\Lambda}(\mathcal{V},\mathcal{X})}} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(X; Y_q|V) - \max_{q' \in \mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(X; Z|V, S)].$$

The corresponding random code secrecy capacity is

$$C^{RC} = \max_{P_X \in \mathcal{P}_\Upsilon(\mathcal{X})} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(X; Y_q) - \max_{q' \in \mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(X; Z|S)].$$

If the main channel $W : \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$ is non-symmetrizable-$\mathcal{X}$, the auxiliary random variable V is not necessary and the stochastic-encoder code secrecy capacity $C^{SC} = C^{RC}$.

The proof is provided in Appendix E. In [25], it is proved that the prefixed channel and the superposition coding scheme are not necessary when the main channel is less noisy than the wiretap channel. However, when considering the stochastic-encoder code capacity of the AVWC with constraints, only the prefixed channel is unnecessary, while the auxiliary random variable V is strictly required to achieve the secrecy capacity in the presence of input and state constraints if the main channel is symmetrizable. In Section VI-B, a numerical example proves that the stochastic-encoder code secrecy capacity of the AVWC with constraints can be strictly smaller than the corresponding random code secrecy capacity.

IV. Proof of Proposition 1

A. Secure Partition Lemma

This subsection presents the secure partition lemma of the AVWC with state constraint. In our previous works [27], [28], Csiszár's almost independent coloring lemma and its extended version were adopted to deal with the strong secrecy problem of the arbitrarily varying wiretap channel and arbitrarily varying multiple access wiretap channel, respectively. In this subsection, we prove that Csiszár's lemma can be used to deal with the strong secrecy problem of the AVWC with constraints.

Define a set of general random variables $\{Y_s\}_{s \in \mathcal{S}}$ depending on $s \in \mathcal{S}$ such that

$$Pr\{Y_s = y\} = P_{Y_s}(y). \tag{14}$$

We write the δ-conditional typical set [19] conditioning on a state sequence $\boldsymbol{s}$ as $T^n_{P_Y,\delta}(\boldsymbol{s})$, and call it typical sequences under the state sequence $\boldsymbol{s}$.

Note that the subscript P_Y in $T^n_{P_Y,\delta}(\boldsymbol{s})$ is a dummy distribution representing the set of distributions $\{P_{Y_s}\}_{s \in \mathcal{S}}$. Note that although the typical set is defined under state sequences, the random variable is not necessarily related to the state. For example, in this paper, all the codewords are uniformly selected at random from a type set $T^n_{P_X}$. Hence, the input random variable X is independent of all the channel states. However, as implied by Lemma 2 in this subsection, the generated codewords are still typical under state sequence with high probability.

Lemma 1 (Asymptotic Equipartition USS): Given $\boldsymbol{x} \in T^n_{P_X,\delta}(\boldsymbol{s})$, for any $\boldsymbol{y} \in T^n_{W,\delta'}[\boldsymbol{x}, \boldsymbol{s}]$, it follows that

$$2^{-n(H_q(Y|X,S)+\epsilon)} \leq W^n_{\boldsymbol{s}}(\boldsymbol{y}|\boldsymbol{x}) \leq 2^{-n(H_q(Y|X,S)-\epsilon)}, \tag{15}$$

where q is the type of sequence $\boldsymbol{s}$, $H_q(Y|X, S) = \sum_{s \in \mathcal{S}} q(s)H(Y_s|X)$, ϵ is a positive real number related to δ and δ' such that $\epsilon \to 0$ as $\delta \to 0$ and $\delta' \to 0$.

The proof is similar to Theorem 1.2 in [29]. By the definition of $H_q(Y|X, S)$, we further define the mutual information $I_q(X; Y|S) = \sum_s q(s)I(X; Y_s)$. Note that the generation of the codewords in this paper and that in [27], [28] is different. In our previous work, all the codewords are i.i.d. generated according to a fixed distribution P_X on $\mathcal{X}$, and in this section, all the codewords are i.i.d. and uniformly selected from a fixed type set $T^n_{P_X}$. However, the codewords are still typical under state sequences with high probability, which is proved in the following lemma.

Lemma 2: Let $\boldsymbol{X}$ be a random sequence uniformly distributed on type set $T^n_{P_X}$. For any given state sequence $\boldsymbol{s} \in \mathcal{S}^n$ and $\delta > 0$, we have

$$Pr\{\boldsymbol{X} \in T^n_{P_X,\delta}(\boldsymbol{s})\} > 1 - 2^{-n\nu} \tag{16}$$

for some $\nu > 0$ and sufficiently large n.

The proof is provided in Appendix B.

Our secure partition lemma is based on Csiszár's almost independent coloring lemma.

Lemma 3 (Lemma 17.3 in [19]): Let P be a distribution on a finite set $\mathcal{V}$ and $\mathcal{F}$ be a subset of $\mathcal{V}$ such that $\mathcal{F} = \{v \in \mathcal{V} : P(v) \leq 1/d\}$. For some positive number ϵ, if $P(\mathcal{F}) \geq 1 - \epsilon$, then a randomly selected mapping $G : \mathcal{V} \to \{1, \ldots, k\}$ satisfies

$$\sum_{m=1}^{k} \left| P(G^{-1}(m)) - \frac{1}{k} \right| \leq 3\epsilon \tag{17}$$

with probability at least $1 - 2ke^{-\epsilon^2(1-\epsilon)d/2k(1+\epsilon)}$.

Moreover, if each P in a family $\mathcal{P}$ satisfies the hypothesis, then the probability that formula (17) holds for all $P \in \mathcal{P}$ is at least $1 - 2k|\mathcal{P}|e^{-\epsilon^2(1-\epsilon)d/2k(1+\epsilon)}$.

Thus, the desired mapping exists if $k \log k < \frac{\epsilon^2(1-\epsilon)d \log e}{2(1+\epsilon) \log 2|\mathcal{P}|}$. This realization of G is denoted by g.

Remark 3: When applying this lemma to the AVWC, the finite set $\mathcal{V}$ represents the codebook $\mathcal{C}$ and $\{1, \ldots, k\}$ represents the message set $\mathcal{M}$. The family of distributions $\mathcal{P}$ is the set of distributions over the codebook $\mathcal{C}$ given all of the sequence pairs $(\boldsymbol{z}, \boldsymbol{s})$. The partition on the codebook $\mathcal{C}$ arises from the mapping $G : \mathcal{C} \to \{1, \ldots, k\}$ constructed in Lemma 3 by considering the disjoint sets $G^{-1}(m), m \in \{1, \ldots, k\}$. Each set $G^{-1}(m) \subseteq \mathcal{C}$ is a sub-codebook corresponding to a message $m \in \{1, \ldots, k\}$. To communicate using the partitioned codebook, in order to transmit message m, the sender uniformly selects a codeword at random from the sub-codebook $G^{-1}(m)$. The partitions on the codebook $\mathcal{C}$ are called 'secure partitions' if the communication using the

partitioned codebooks under any $\boldsymbol{s} \in \mathcal{S}^n$ achieves arbitrarily small information leakage $I(M; \boldsymbol{Z_s})$.

The secure partition is constructed on 'good' codebooks, which is defined in the following definition.

Definition 9 (Definition of 'Good' Codebook): Let $\mathcal{C}$ be a codebook containing $\widetilde{N} = 2^{n\widetilde{R}}$ codewords with type P_X for some $\widetilde{R} > 0$. For $\delta > 0$ and $\nu > 0$, the codebook $\mathcal{C}$ is called a 'good' codebook if

$$|T^n(\mathcal{C}, \boldsymbol{s})| > (1 - 2 \cdot 2^{-n\nu}) 2^{n\widetilde{R}} \text{ for all } \boldsymbol{s} \in \mathcal{S}^n, \tag{18}$$

where $T^n(\mathcal{C}, \boldsymbol{s}) = \mathcal{C} \cap T^n_{P_X, \delta}(\boldsymbol{s})$ is the set of codewords that is typical under state sequence $\boldsymbol{s}$ in codebook $\mathcal{C}$.

The following lemma proves that the codebooks generated by random selection from a fixed type set is 'good' with high probability.

Lemma 4 (The Existence of 'Good' Codebook): Let $\boldsymbol{C} = \{\boldsymbol{X}(1), \ldots, \boldsymbol{X}(\widetilde{N})\}$ be a codebook containing $\widetilde{N} = 2^{n\widetilde{R}}$ codewords for some $\widetilde{R} > 0$, each uniformly distributed on a fixed type set $T^n_{P_X}$. Then the probability of codebook $\boldsymbol{C}$ being 'good' is bounded by

$$Pr\{\boldsymbol{C} \text{ is 'good'}\} > 1 - \zeta, \tag{19}$$

where ζ is a double exponentially small number and $\zeta \to 0$ as $n \to \infty$.

With the help of Lemma 2 and the Chernoff bound, the proof of the above lemma is almost the same as that in [27] [28, Appendix C]. The following lemma proves the existence of the secure partitions on 'good' codebooks used for the AVWC with constraints.

Lemma 5 (Secure Partition Lemma): Let τ, ε be two positive real numbers that can be arbitrarily small. Let $\mathcal{E} = \{E_s : s \in \mathcal{S}\}$ be an AVC with input alphabet $\mathcal{X}$, state alphabet $\mathcal{S}$ and output alphabet $\mathcal{Z}$ and $\mathcal{M}$ be the message set. Furthermore, let $\mathcal{C} = \{\boldsymbol{x}(i) : 1 \le i \le \widetilde{N}\}$ be a 'good' codebook for $\widetilde{N} = 2^{n\widetilde{R}}$ and $\widetilde{R} > 0$. Suppose $\boldsymbol{Z_s}$ is the channel output with input sequence $\boldsymbol{X}$ uniformly distributed on the codebook $\mathcal{C}$ under state sequence $\boldsymbol{s}$. For real number $N' = 2^{nR'}$ satisfying

$$R' > \max_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I_q(X; Z|S)$$

and $R = \widetilde{R} - R'$, where $P_{X,S,Z}(x, s, z) = P_X(x) q(s) E(z|x, s)$, there exists a secure partition g on codebook $\mathcal{C}$ which divides the codebook into $N = 2^{n(R-\tau)}$ sub-codebooks $\{\mathcal{C}(m)\}_{m \in \{1, \ldots, N\}}$. Each sub-codebook corresponds to a message m. In the communication using the partitioned codebook, the sender uniformly selects a codeword $\boldsymbol{X}$ at random from sub-codebook $\mathcal{C}(M)$ to transmit message M. Consequently,

$$I(M; \boldsymbol{Z_s}) < \varepsilon, \quad \boldsymbol{s} \in \mathcal{S}^n(\Lambda)$$

for some $\varepsilon \to 0$ as $n \to \infty$, where n is the length of sequences depending on τ and ε, M is the index of the sub-codebook $\mathcal{C}(M)$ containing the codeword $\boldsymbol{X}$.

Remark 4: For the AVWC free of constraints, one can enlarge $I_q(X; Z|S) = \sum_{s \in \mathcal{S}} q(s) I(X; Z_s)$ to $\max_s I(X; Z_s)$. In this case, q is a one point distribution satisfying $q(s^*) = 1$ for $s^* = \arg\max_{s \in \mathcal{S}} I(X; Y_s)$ and $q(s) = 0$ otherwise, and the corresponding state sequence $\boldsymbol{s}$ is the symbol s^* repeated n times. However, this does not work in the presence of the state constraint since such a state sequence violates the state constraint if $l(s^*) > \Lambda$.

Proof of Lemma 5: Let $(\boldsymbol{X}, \boldsymbol{Z_s})$ be a pair of random sequences as in Lemma 5 and $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$. By Lemma 2.12 in [19], the definition of 'good' codebook and Hoeffding's inequality, it follows that

$$Pr\{(\boldsymbol{X}, \boldsymbol{Z_s}) \in T^n_{P_{XZ}, 2\delta}(\boldsymbol{s})\} \ge 1 - 2^{-n\nu} \tag{20}$$

for some $\delta > 0$, where $\nu \to 0$ as $\delta \to 0$ and $n \to \infty$. For any $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$, we construct a high probability subset $\mathcal{B}(\boldsymbol{s}) \subseteq \mathcal{Z}^n$ with the help of the following two auxiliary sets.

$$\mathcal{B}_0(\boldsymbol{s}) = \left\{\boldsymbol{z} \in T^n_{P_Z, \delta'}(\boldsymbol{s}) : \Psi(\boldsymbol{z}, \boldsymbol{s}) < 2^{-\frac{n\nu}{2}}\right\} \tag{21}$$

and

$$\mathcal{B}_1(\boldsymbol{s}) = \left\{\boldsymbol{z} : Pr\{\boldsymbol{Z_s} = \boldsymbol{z}\} < 2^{-n\frac{\nu}{2}} \prod_{i=1}^n P_{Z_{s_i}}(z_i)\right\}, \tag{22}$$

where $\Psi(\boldsymbol{z}, \boldsymbol{s}) = Pr\{\boldsymbol{X} \notin T^n_{P_{X|Z}, 2\delta}[\boldsymbol{z}, \boldsymbol{s}] | \boldsymbol{Z_s} = \boldsymbol{z}\}$, $\delta' > 0$ depends on δ, and the subscript P_Z in (21) is a dummy distribution such that for any $s \in \mathcal{S}$,

$$P_{Z_s}(\boldsymbol{z}) = \sum_x P_X(x) E(z|x, s).$$

Now by formula (20), we have

$$\begin{aligned}
&1 - 2^{-n\nu} \\
&\le \sum_{\boldsymbol{z} \in \mathcal{B}_0(\boldsymbol{s})} Pr\{\boldsymbol{Z_s} = \boldsymbol{z}\} Pr\{\boldsymbol{X} \in T^n_{P_{X|Z}, 2\delta}[\boldsymbol{z}, \boldsymbol{s}] | \boldsymbol{Z_s} = \boldsymbol{z}\} \\
&\quad + \sum_{\boldsymbol{z} \in T^n_{P_Z, \delta'}(\boldsymbol{s}) \backslash \mathcal{B}_0(\boldsymbol{s})} Pr\{\boldsymbol{Z_s} = \boldsymbol{z}\} \\
&\quad \times Pr\left\{\boldsymbol{X} \in T^n_{P_{X|Z}, 2\delta}[\boldsymbol{z}, \boldsymbol{s}] | \boldsymbol{Z_s} = \boldsymbol{z}\right\} \\
&\overset{(a)}{<} Pr\{\boldsymbol{Z_s} \in \mathcal{B}_0(\boldsymbol{s})\} + (1 - Pr\{\boldsymbol{Z_s} \in \mathcal{B}_0(\boldsymbol{s})\})(1 - 2^{-\frac{n\nu}{2}}),
\end{aligned}$$

where (a) follows by the definition of $\mathcal{B}_0(\boldsymbol{s})$ and $\Psi(\boldsymbol{z}, \boldsymbol{s})$ in (21). Hence,

$$Pr\{\boldsymbol{Z_s} \in \mathcal{B}_0(\boldsymbol{s})\} > 1 - 2^{-\frac{n\nu}{2}}. \tag{23}$$

For $\mathcal{B}_1(\boldsymbol{s})$, it follows that

$$\begin{aligned}
Pr\{\boldsymbol{Z} \in \mathcal{B}_1(\boldsymbol{s})\} &= \sum_{\boldsymbol{z} \in \mathcal{B}_1(\boldsymbol{s})} Pr\{\boldsymbol{Z_s} = \boldsymbol{z}\} \\
&< \sum_{\boldsymbol{z} \in \mathcal{B}_1(\boldsymbol{s})} 2^{-n\frac{\nu}{2}} \prod_{i=1}^n P_{Z_{s_i}}(z_i) \le 2^{-n\frac{\nu}{2}}.
\end{aligned} \tag{24}$$

Setting $\mathcal{B}(\boldsymbol{s}) = \mathcal{B}_0(\boldsymbol{s}) \backslash \mathcal{B}_1(\boldsymbol{s})$ yields

$$Pr\{\boldsymbol{Z_s} \in \mathcal{B}(\boldsymbol{s})\} > 1 - 2 \cdot 2^{-\frac{n\nu}{2}}. \tag{25}$$

Then we set parameters for Lemma 3 as follows:

$$\begin{aligned}
&\epsilon = 2^{-\frac{n\nu}{2}}, d = 2^{n(R - \frac{\tau}{2})}, \\
&k = 2^{n(R-\tau)}, \\
&\mathcal{P} = \{P^n_{\boldsymbol{X}}\} \cup \{P^n_{\boldsymbol{X}|\boldsymbol{z},\boldsymbol{s}} : \boldsymbol{s} \in \mathcal{S}^n(\Lambda), \boldsymbol{z} \in \mathcal{B}(\boldsymbol{s})\},
\end{aligned} \tag{26}$$

where $P^n_{\boldsymbol{X}}$ and $P^n_{\boldsymbol{X}|\boldsymbol{z},\boldsymbol{s}}$ are distributions on codebook $\mathcal{C}$. The verification of the parameters is given in Appendix C. Note that when applying Lemma 3 to our model, the finite set $\mathcal{V}$ is the codebook $\mathcal{C}$ and the integer set $[1,\ldots,k]$ is the message set $\mathcal{M}$. Now let $g:\mathcal{C}\to\mathcal{M}$ be a realization of the random mapping G satisfying formula (17) in Lemma 3, by the definition of $\mathcal{P}$ in (26) and formula (17), we have

$$\sum\nolimits_{m=1}^{k}\left|Pr\{M_g(\boldsymbol{X})=m\}-\frac{1}{k}\right|<3\epsilon, \tag{27}$$

$$\sum\nolimits_{m=1}^{k_1}\left|Pr\{M_g(\boldsymbol{X})=m|\boldsymbol{Z}_{\boldsymbol{s}}=\boldsymbol{z}\}-\frac{1}{k}\right|<3\epsilon, \tag{28}$$

where $M_g(\boldsymbol{X})$=$g(\boldsymbol{X})$. The term $M_g(\boldsymbol{X})$ is a random variable distributed on the message set, and its distribution is defined by the mapping g. By formula (27), M_g is almost uniformly distributed on the message set. Applying the uniform continuity of entropy (Lemma 2.7 in [19]) to (28) yields

$$|H(M_g(\boldsymbol{X})|\boldsymbol{Z}_{\boldsymbol{s}}=\boldsymbol{z})-H(M)|\le -3\epsilon\log\frac{3\epsilon}{k}.$$

Combining the above inequality and formula (25) gives

$$H(M_g(\boldsymbol{X})|\boldsymbol{Z}_{\boldsymbol{s}})\ge(1-2\epsilon)(\log k+3\epsilon\log\frac{3\epsilon}{k})$$

and

$$I(M_g(\boldsymbol{X});\boldsymbol{Z}_{\boldsymbol{s}})\le 5\epsilon\log k-3\epsilon\log 3\epsilon \tag{29}$$

for all $\boldsymbol{s}\in\mathcal{S}^n(\Lambda)$. Note that the partition g is not necessarily equally divided. To construct an equal partition, the following lemma is needed.

Lemma 6 (Lemma 4 in [30]): For any given codebook $\mathcal{C}$, if the function $g:\mathcal{C}\to[1:k]$ satisfies (27), there exists a partition $\{\mathcal{C}_m\}_{m=1}^k$on $\mathcal{C}$ such that

1) $|\mathcal{C}_m|=\frac{2^{n\tilde{R}}}{k}$ for all $m\in[1:k]$,
2) $H(M|M_g(\boldsymbol{X}))<4\sqrt{\epsilon}\log k$,

where M is the index of the sub-codebook containing $\boldsymbol{X}$.

Considering the above Lemma 6 and (29), the information leakage is upper bounded by

$$\begin{aligned}I(M;\boldsymbol{Z}_{\boldsymbol{s}})&\le I(M,M_g(\boldsymbol{X});\boldsymbol{Z}_{\boldsymbol{s}})\\&\le I(M_g(\boldsymbol{X});\boldsymbol{Z}_{\boldsymbol{s}})+I(M;\boldsymbol{Z}_{\boldsymbol{s}}|M_g(\boldsymbol{X}))\\&\le 5\epsilon\log k-3\epsilon\log 3\epsilon+4\sqrt{\epsilon}\log k\le\varepsilon\end{aligned}$$

for all $\boldsymbol{s}\in\mathcal{S}(\Lambda)$. Since $\epsilon=2^{-n\frac{\nu}{2}}$, the information leakage ε can be arbitrarily small as $n\to\infty$. □

B. Coding Scheme

To prove our first main result, we use a coding scheme similar to that in [8], except that it is a stochastic-encoder code version. The details of the stochastic coding scheme are presented as follows.

1) Codebook Generation: Let $\mathcal{M}$ be the message set. For a fixed type P_X on $\mathcal{X}$ with $\min_x P_X(x)>0$ and $\psi(P_X)\le\Upsilon,\Lambda_0(P_X)\ge\Lambda+\alpha,\alpha>0$, uniformly and i.i.d. select $\widetilde{N}$ codewords from type set $T^n_{P_X}$. The set of codewords is denoted by codebook $\mathcal{C}=\{\boldsymbol{x}(i):1\le i\le\widetilde{N}\}$. Then the codebook is divided into $N=2^{n(R-\tau)}$ sub-codebooks for $\tau>0$, each with $N'=2^{nR'}$ codewords, where $R'=\widetilde{R}-R$. The m^{th} sub-codebook is denoted by $\mathcal{C}(m)$ and the l^{th} codeword in sub-codebook $\mathcal{C}(m)$ is denoted by $\boldsymbol{x}(m,l)$.

2) Encoding: To transmit message m, the encoder uniformly selects a codeword $\boldsymbol{x}(m,l)$ at random from sub-codebook $\mathcal{C}(m)$.

3) Decoding: Let (X,S,Y) be dummy random variables distributed according to P_{XSY}, which is the joint type of $(\boldsymbol{x}(m,l),\boldsymbol{s},\boldsymbol{y})$. Let $D(\cdot||\cdot)$ be the KL divergence. The received sequence $\boldsymbol{y}$ belongs to the decoding set $\mathcal{D}_{m,l}$ if and only if

1) the joint type P_{XSY} satisfies $D(P_{XSY}||P_X\times P_S\times W)\le\eta$,
2) for each $\boldsymbol{x}(m',l')$ with $(m',l')\ne(m,l)$ such that $D(P_{X'SY}||P_{X'}\times P_S\times W)\le\eta$ for some $\boldsymbol{s}'\in\mathcal{S}^n$ satisfying $l^n(\boldsymbol{s}')\le\Lambda$, we have $I(X,Y;X'|S)\le\eta$. Here X,X',S,S',Y are dummy random variables distributed according to $P_{XX'SY}$, which is the joint type of $(\boldsymbol{x}(m,l),\boldsymbol{x}(m',l'),\boldsymbol{s},\boldsymbol{s}',\boldsymbol{y})$.

Remark 5: Note that the decoding rule in the above coding scheme is the same as that in [8]. This is because we adopt a strict decoding rule for the stochastic-encoder code. The decoder is required to decode the pair (m,l), not just the message m. Thus, the decoding procedure is the same as the deterministic code version in [8].

Now by the analysis used in [8], the legitimate receiver can decode the message correctly if

$$\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q)-\tau\le\widetilde{R}\le\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q)-\frac{2}{3}\tau \tag{30}$$

for some $\tau>0$ that can be arbitrarily small with codeword type satisfying $\Lambda_0(P_X)\ge\Lambda+\alpha,\alpha>0$. Note that any codebook with codeword type $\Lambda_0(P_X)<\Lambda$ cannot be used in reliable communication. By Lemma 4, a codebook that is generated as in the above stochastic coding scheme is a 'good codebook' with probability of at least $1-\zeta$, where ζ is a double exponentially small number. Hence, we assume our deterministic codebook is a 'good codebook'. Lemma 5 implies that the transmission achieves strong secrecy if $R'>\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I_q(X;Z|S)$ and the reliable and secure communication can be achieved if

$$\begin{aligned}&\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q)-\max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})}I_{q'}(X;Z|S)-2\tau\\&\quad\le\frac{1}{n}\log|\mathcal{M}|=R\\&\le\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q)-\max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})}I_{q'}(X;Z|S)-\frac{5}{3}\tau.\end{aligned}$$

By the continuity of mutual information and a similar argument as that in the proof of Theorem 3 in [8], the case that $\alpha=0$ can be asymptotically achieved.

C. Prefixed Channel

The prefixed channel $P_{X|U}$ is a discrete memoryless channel (DMC) prefixed to the AVWC $(\mathcal{W},\mathcal{E})$, which is selected by the sender and will not be affected by the jammer. The concatenated channel is controlled by the sender and the jammer separately and the jammer can 'see' the prefixed channel selection of the sender. Hence, the prefixed channel can be interpreted as a fixed state DMC with channel state s_p.

The state space of the concatenated channel is $\mathcal{S}^* = s_p \times \mathcal{S}$. Since the state s_p is fixed by the sender, we assume that $l(s_p \times s) = l(s)$ for any $s^* = (s_p \times s) \in \mathcal{S}^*$.

The main channel with a prefixed channel is defined by

$$\begin{aligned} W(y|u, s^*) &= \sum_x W(y|x, s^*) P_{X|U}(x|u) \\ &= \sum_x W(y|x, s) P_{X|U}(x|u) \end{aligned}$$

for any $(u, x, s^*, y) \in \mathcal{U} \times \mathcal{X} \times \mathcal{S}^* \times \mathcal{Y}$ and the average channel is

$$W_{q^*}(y|u) = \sum_s W(y|u, s^*) q^*(s^*)$$

for $q^* \in \mathcal{P}(\mathcal{S}^*)$. For such a concatenated channel, by previous subsections, it follows that all numbers R satisfying

$$R \le \min_{q^* \in \mathcal{P}_\Lambda(\mathcal{S}^*)} I(U; Y_{q^*}) - \max_{q^{*\prime} \in \mathcal{P}_\Lambda(\mathcal{S}^*)} I_{q^{*\prime}}(U; Z|S^*)$$

is achievable. Note that the distribution $q^* \in \mathcal{P}(\mathcal{S}^*)$ is in fact $q \in \mathcal{P}(\mathcal{S})$ since s_p is fixed. As a result, for given distribution $q^* \in \mathcal{P}(\mathcal{S}^*)$, the average channel is in fact specified by

$$\begin{aligned} W_{q^*}(y|u) &= \sum_{s^* \in \mathcal{S}^*} W(y|u, s^*) q^*(s^*) \\ &= \sum_{s \in \mathcal{S}} \sum_x W(y|x, s) P_{X|U}(x|u) q(s) \\ &= \sum_x P_{X|U}(x|u) \sum_{s \in \mathcal{S}} W(y|x, s) q(s) \\ &= \sum_x P_{X|U}(x|u) W_q(y|x) = W_q(y|u). \end{aligned}$$

For state constraint, it follows that $\mathbb{E}[l(S^*)] = \sum_{s^* \in \mathcal{S}^*} l(s^*) q^*(s^*) = \sum_{s \in \mathcal{S}} l(s) q(s) = \mathbb{E}[l(S)]$ and the minimum or maximum taken over $\mathcal{P}_\Lambda(\mathcal{S}^*)$ is in fact taken over $\mathcal{P}_\Lambda(\mathcal{S})$.

For input distribution, as we discussed in Remark 2, the AVWC may be symmetrizable-$\mathcal{U}$ after adding a prefixed channel even if the original channel is non-symmetrizable, which leads to $\Lambda_1(P_{UX}) \le \Lambda_0(P_X)$ for joint distribution P_{UX} and P_X such that $P_X(x) = \sum_u P_{UX}(u, x)$. Maximizing over input distributions satisfying $\Lambda_1(P_{UX}) \ge \Lambda$ and $\psi(P_{UX}) \le \Upsilon$, the proof is completed. □

V. Proof of Theorem 5

This section proves the lower bound of the secrecy capacity of the AVWC with constraints using superposition stochastic-encoder code in Theorem 5. The secure partition lemma is given in Subsection V-A and the achievability is proved in Subsection V-B.

A. Secure Partition Lemma for Superposition Stochastic Coding Scheme

To prove the secure partition lemma for the superposition coding scheme, we first extend the definition of 'good codebooks' as follows.

Definition 10 ('Good Codebooks' for Superposition Coding): Let $N_C = 2^{nR_C}$ and $\widetilde{N} = 2^{n\widetilde{R}}$ be positive integers with $R_C > 0$ and $\widetilde{R} > 0$. For superposition coding scheme, let $\mathcal{C}_C = \{\boldsymbol{v}(i), 1 \le i \le N_C\}$ be the cloud center 'good' codebook satisfying Definition 9 with size $|\mathcal{C}_C| = N_C$. Let $\{\mathcal{C}(i), 1 \le i \le N_C\}$ be a set of codebooks with each element $\mathcal{C}(i)$ corresponding to a sequence $\boldsymbol{v}(i)$ and $|\mathcal{C}(i)| = \widetilde{N}$. The superposition codebooks $(\mathcal{C}_C, \{\mathcal{C}(i), 1 \le i \le N_C\})$ are 'good superposition codebooks' if for $\delta > 0$ and $\nu > 0$,

$$\begin{aligned} |T^n(\mathcal{C}(i), \boldsymbol{v}(i), \boldsymbol{s})| &> (1 - 2 \cdot 2^{-n\nu}) 2^{n\widetilde{R}} \\ &\text{for all } \boldsymbol{s} \in \mathcal{S}^n(\Lambda) \text{ and } \boldsymbol{v}(i) \in (\mathcal{C}_C \cap T^n_{P_V, \delta}(\boldsymbol{s})), \end{aligned}$$

where $T^n(\mathcal{C}(i), \boldsymbol{v}(i), \boldsymbol{s}) = \{\boldsymbol{x}(i, j) \in \mathcal{C}(i) : (\boldsymbol{v}(i), \boldsymbol{x}(i, j)) \in T^n_{P_{VX}, 2\delta}(\boldsymbol{s})\}$, where $P_{VX}(v, x) = P_V(v) P_{X|V}(x|v)$ for $v \in \mathcal{V}, x \in \mathcal{X}$. Then codebooks $\{\mathcal{C}(i)\}_{i=1}^{N_C}$ are said to be 'good' codebooks w.r.t. $\mathcal{C}_C$.

Remark 6: Since a randomly selected sequence from a type set is typical under a given state sequence $\boldsymbol{s}$ with high probability, which is proved in Lemma 2, it is sufficient to generate $2^{n\eta}$ codewords for $\mathcal{C}_C$ with some $\eta > 0$.

Lemma 7: Let $\boldsymbol{C}_C = \{\boldsymbol{V}(1), \ldots, \boldsymbol{V}(N_C)\}$ be a random codebook containing N_C codewords, each uniformly distributed on a type set $T^n_{P_V}$. Let $\boldsymbol{C}(i) = \{\boldsymbol{X}(i, j), 1 \le j \le \widetilde{N}\}, 1 \le i \le N_C$ such that

$$Pr\{\boldsymbol{X}(i, j) = \boldsymbol{x} | \boldsymbol{V}(i) = \boldsymbol{v}\} = 1/|T^n_{P_{X|V}}[\boldsymbol{v}]| \text{ for } \boldsymbol{x} \in T^n_{P_{X|V}}[\boldsymbol{v}]$$

and $Pr\{\boldsymbol{X}(i, j) = \boldsymbol{x} | \boldsymbol{V}(i) = \boldsymbol{v}\} = 0$ otherwise. The probability that $(\boldsymbol{C}_C, \{\boldsymbol{C}(i), 1 \le i \le N_C\})$ are 'good superposition codebooks' is bounded by

$$\begin{aligned} Pr\{(\boldsymbol{C}_C, \{\boldsymbol{C}(i), 1 \le i \le N_C\}) \text{ are 'good} \\ \text{superposition codebooks'}\} > 1 - \zeta, \end{aligned}$$

where ζ is a double exponentially small number and $\zeta \to 0$ as $n \to \infty$.

Proof: By Lemma 4, we have

$$Pr\{\boldsymbol{C}_C \text{ is a 'good codebook'}\} > 1 - \zeta_1$$

for some double exponentially small number $\zeta_1 \to 0$ as $n \to \infty$. Now suppose $\boldsymbol{C}_C$ is a 'good' codebook and $|\boldsymbol{C}_C \cap T^n_{P_V, \delta}(\boldsymbol{s})| \le 2^{nR_C}$. By Lemma 2 and the property of the 'good' codebook, for any $\boldsymbol{V}(i) \in (\boldsymbol{C}_C \cap T^n_{P_V, \delta}(\boldsymbol{s}))$ and $\boldsymbol{X} \in \boldsymbol{C}(i)$, it follows that

$$Pr\{(\boldsymbol{V}(i), \boldsymbol{X}) \in T^n_{P_{VX}, 2\delta}(\boldsymbol{s})\} > 1 - 2^{-n\nu}$$

for some $\nu > 0$ tending to a constant related to δ as $n \to \infty$. Applying the same technique used in Lemma 4, we have

$$Pr\{|T^n(\boldsymbol{C}(i), \boldsymbol{V}(i), \boldsymbol{s})| > (1 - 2 \cdot 2^{-n\nu}) 2^{n\widetilde{R}}\} > 1 - \zeta_2$$

for some double exponentially small number $\zeta_2 \to 0$ as $n \to \infty$. Then the probability that $(\boldsymbol{C}_C, \{\boldsymbol{C}(i), 1 \le i \le N_C\})$ are 'good superposition codebooks' is bounded by

$$\begin{aligned} &Pr\{(\boldsymbol{C}_C, \{\boldsymbol{C}(i), 1 \le i \le N_C\}) \\ &\quad \text{are 'good superposition codebooks'}\} \\ &\ge (1 - \zeta_1) \cdot (1 - \zeta_2)^{2^{nR_C}} \overset{(a)}{>} (1 - \zeta_1) \cdot (1 - 2^{nR_C} \zeta_2) \overset{(b)}{>} 1 - \zeta \end{aligned}$$

for ζ depending on ζ_1 and ζ_2 and $\zeta \to 0$ as $n \to \infty$, where (a) follows by Bernoulli's inequality and (b) follows from the

fact that ζ_1 and ζ_2 are double exponentially small numbers. The proof is completed. □

The following lemma deals with the secure transmission using the superposition stochastic coding scheme.

Lemma 8 (Secure Partition Lemma for Superposition Coding): Let τ, ε be positive real numbers and $\mathcal{E} = \{E_s : s \in \mathcal{S}\}$ be an AVC with auxiliary alphabet $\mathcal{V}$, input alphabet $\mathcal{X}$, state alphabet $\mathcal{S}$ and output alphabet $\mathcal{Z}$ such that

$$E^n(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}) = \prod_{i=1}^{n} E(z_i|x_i, s_i) = \prod_{i=1}^{n} E_{s_i}(z_i|x_i),$$

where $\boldsymbol{x} \in \mathcal{X}^n, \boldsymbol{s} \in \mathcal{S}^n, \boldsymbol{z} \in \mathcal{Z}^n$. Furthermore, let $\mathcal{C}_C = \{\boldsymbol{v}(i)\}_{i=1}^{N_C}$ be the 'cloud center' good codebook with codewords uniformly selected from a type set $T^n_{P_V}$, where $N_C = 2^{n\eta}$ and $\eta > 0$. For each 'cloud center' codeword $\boldsymbol{v}(i)$ typical under given state sequence $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$, let $\mathcal{C}(i) = \{\boldsymbol{x}(i,j)\}_{j=1}^{\widetilde{N}}$ be a 'good' codebook with each codeword i.i.d. and uniformly selected from the conditional type set $T^n_{P_{X|V}}[\boldsymbol{v}(i)]$, where $\widetilde{N} = 2^{n\widetilde{R}}$ and $\widetilde{R} > 0$. Suppose $\boldsymbol{Z_s}$ is the channel output with input sequence $\boldsymbol{X}$ under state sequence $\boldsymbol{s}$. For real number $N' = 2^{nR'}$ satisfying

$$R' > \max_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I_q(X; Z|S, V),$$

and $R = \widetilde{R} - R'$, where $P_{VXSZ}(v, x, s, z) = P_V(v)P_{X|V}(x|v)q(s)E(z|x, s)$, there exists a secure partition g_i on codebook $\mathcal{C}(i)$, which divides the codebook into $N = 2^{n(R-\tau)}$ sub-codebooks $\{\mathcal{C}(i, m)\}_{m=1}^{N}$. To transmit message M using the partitioned codebooks, the sender unformly selects a 'cloud center' codeword index i at random and then a codeword $\boldsymbol{X}$ from the sub-codebook $\mathcal{C}(i, M)$. Consequently,

$$I(M; \boldsymbol{Z_s}) < \varepsilon$$

for any $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and arbitrarily small $\varepsilon > 0$.

Proof: The proof is similar to that of Lemma 5, with only minor modification of the parameters required. Specifically, for given state sequence $\boldsymbol{s}$ and $\boldsymbol{v}(i)$ typical under $\boldsymbol{s}$, the auxiliary sets $\mathcal{B}_0(\boldsymbol{s})$ and $\mathcal{B}_1(\boldsymbol{s})$ are rewritten as

$$\mathcal{B}_0(\boldsymbol{s}) = \left\{(\boldsymbol{v}(i), \boldsymbol{z}) \in T^n_{P_{VZ}, 2\delta}(\boldsymbol{s}) : \Psi(\boldsymbol{v}(i), \boldsymbol{z}, \boldsymbol{s}) < 2^{-\frac{n\nu}{2}}\right\} \tag{31}$$

and

$$\mathcal{B}_1(\boldsymbol{s}) = \{(\boldsymbol{v}(i), \boldsymbol{z}) : Pr\{\boldsymbol{Z_s} = \boldsymbol{z}|\boldsymbol{V} = \boldsymbol{v}(i)\} < 2^{-n\frac{\nu}{2}} \prod_{k=1}^{n} P_{Z_{s_k}|V}(z_k|v_k)\}, \tag{32}$$

where $\Psi(\boldsymbol{v}(i), \boldsymbol{z}, \boldsymbol{s}) = Pr\{\boldsymbol{X} \notin T^n_{P_{X|VZ}, 3\delta}[\boldsymbol{v}(i), \boldsymbol{z}, \boldsymbol{s}]|\boldsymbol{V} = \boldsymbol{v}(i), \boldsymbol{Z_s} = \boldsymbol{z}\}$ and P_{VZ} is a dummy distribution such that for any $s \in \mathcal{S}$, $P_{VZ_s}(v, z) = \sum_x P_V(v)P_{X|V}(x|v)E(z|x, s)$.

The parameters in (26) are modified as

$$\begin{aligned}
&\epsilon = 2^{-\frac{n\nu}{2}}, d = 2^{n(R-\frac{\tau}{2})},\\
&k = 2^{n(R-\tau)},\\
&\mathcal{P} = \{P^n_{\boldsymbol{X}|\boldsymbol{v}(i)}\} \cup\\
&\{P^n_{\boldsymbol{X}|\boldsymbol{v}(i),\boldsymbol{z},\boldsymbol{s}} : \boldsymbol{v}(i) \in (\mathcal{C}_C \cap T^n_{P_V}(\boldsymbol{s})), \boldsymbol{z} \in \mathcal{B}(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{S}^n(\Lambda)\},
\end{aligned} \tag{33}$$

where $P^n_{\boldsymbol{X}|\boldsymbol{v}(i)}$ and $P^n_{\boldsymbol{X}|\boldsymbol{v}(i),\boldsymbol{z},\boldsymbol{s}}$ are distributions on codebook $\mathcal{C}(i)$. The verification of the parameters is also provided in Appendix C. The main difference between the parameters here and those in Lemma 5 is that the distributions here are conditioning on the 'cloud center' $\boldsymbol{v}(i)$, which means the eavesdropper has the knowledge of $\boldsymbol{v}(i)$. This is reasonable since we do not assume anything about the wiretap channel. Now by the definition of $\mathcal{P}$ in (33) and formula (17), we have

$$\sum\nolimits_{m=1}^{k} \left| Pr\{M_{g_i}(\boldsymbol{X}) = m|\boldsymbol{V} = \boldsymbol{v}(i)\} - \frac{1}{k} \right| < 3\epsilon, \tag{34}$$

$$\sum\nolimits_{m=1}^{k} \left| Pr\{M_{g_i}(\boldsymbol{X}) = m|\boldsymbol{Z_s} = \boldsymbol{z}, \boldsymbol{V} = \boldsymbol{v}(i)\} - \frac{1}{k} \right| < 3\epsilon, \tag{35}$$

where $M_{g_i}(\boldsymbol{X}) = g_i(\boldsymbol{X})$. Applying the uniform continuity of entropy (Lemma 2.7 in [19]) to (35) yields

$$|H(M_{g_i}(\boldsymbol{X})|\boldsymbol{Z_s} = \boldsymbol{z}, \boldsymbol{V} = \boldsymbol{v}(i)) - H(M)| \le -3\epsilon \log \frac{3\epsilon}{k}.$$

Since $\mathcal{C}_C$ is a 'good' codebook, combining the above inequality and formula (25) together with the fact that $\boldsymbol{v}(i)$ is typical under $\boldsymbol{s}$ with high probability in Definition 9 gives

$$\begin{aligned}
H(M_{g_i}(\boldsymbol{X})|\boldsymbol{Z_s}) &\ge H(M_{g_i}(\boldsymbol{X})|\boldsymbol{Z_s}, \boldsymbol{V})\\
&\ge (1 - 2\epsilon)(1 - 2\epsilon_1)(\log k + 3\epsilon \log \frac{3\epsilon}{k})\\
&\overset{(a)}{\ge} (1 - 4\epsilon_2)(\log k + 3\epsilon \log \frac{3\epsilon}{k})
\end{aligned}$$

and

$$I(M_{g_i}(\boldsymbol{X}); \boldsymbol{Z_s}) \le (3\epsilon + 4\epsilon_2) \log k - 3\epsilon \log 3\epsilon \tag{36}$$

for all $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$, where (a) follows by setting $\epsilon_2 = \max\{\epsilon, \epsilon_1\}$. Applying Lemma 6 again yields

$$I(M; \boldsymbol{Z_s}) \le \varepsilon$$

for all $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and arbitrarily small ε. The proof is completed. □

B. Coding Scheme

The following coding scheme is a stochastic superposition coding scheme used over an AVWC with input constraint Υ and state constraint Λ originated from Definition 11 in [20]. We first set $U = X$ and then apply the prefixed channel discussed in Section IV-C.

1) Codebook Generation: Let $\mathcal{M}$ be the message set. For a fixed type P_V on $\mathcal{V}$ with $\min_{v \in \mathcal{V}} P_V(v) \ge \beta > 0$, i.i.d. and unformly select $N_C = 2^{nR_C}$ codewords from the type set $T^n_{P_V}$ for some $R_C > 0$. The set of codewords is denoted by codebook $\mathcal{C}_C = \{\boldsymbol{v}(i)\}_{i=1}^{N_C}$. For each sequence $\boldsymbol{v}(i)$, i.i.d. and uniformly select $\widetilde{N} = 2^{N\widetilde{R}}$ codewords for $\widetilde{R} > 0$ from the set $T^n_{P_{X|V}}[\boldsymbol{v}(i)]$ with $\min_{x,v} P_{X|V}(x|v) > 0$ and $\psi(P_{VX}) \le \Upsilon, \Lambda_0(P_{VX}) \ge \Lambda + \alpha, \alpha > 0$. The set of codewords is denoted by codebook $\mathcal{C}(i) = \{\boldsymbol{x}(i,j)\}_{j=1}^{\widetilde{N}}$. Then each codebook is divided into $N = 2^{nR}$ sub-codebooks by a partition constructed in Lemma 8, each with $N' = 2^{nR'}$ codewords, where $R > 0, R' > 0, R' = \widetilde{R} - R$. The m^{th} sub-codebook of $\mathcal{C}(i)$ is denoted by $\mathcal{C}(i, m)$ and the l^{th} codeword in sub-codebook $\mathcal{C}(i, m)$ is denoted by $\boldsymbol{x}(i, m, l)$.

2) Encoding: To transmit message m, the encoder uniformly selects a common message codeword $\boldsymbol{v}(i)$ at random and codeword $\boldsymbol{x}(i,m,l)$ from sub-codebook $\mathcal{C}(i,m)$.

3) Decoding: Let (V,X,S,Y) be random variables distributed according to the joint type of $(\boldsymbol{v}(i),\boldsymbol{x}(i,m,l),\boldsymbol{s},\boldsymbol{y})$. The received sequence $\boldsymbol{y}$ belongs to the decoding set $\mathcal{D}_{i,m,l}$ if and only if

1) the joint type P_{VXSY} satisfies $D(P_{VXSY}||P_{VX}\times P_S\times W)\leq\eta$,
2) for each $(\boldsymbol{v}(i'),\boldsymbol{x}(i',m',l'))$ with $(i',m',l')\neq(i,m,l)$ such that $D(P_{V'X'S'Y}||P_{V'X'}\times P_{S'}\times W)\leq\eta$ for some $\boldsymbol{s}'\in\mathcal{S}^n$ satisfying $l^n(\boldsymbol{s})\leq\Lambda$, we have $I(V,X,Y;V',X'|S)\leq\eta$. Here, (V,V',X,X',S,S',Y) are dummy random variables distributed according to the joint type of $(\boldsymbol{v}(i),\boldsymbol{v}(i'),\boldsymbol{x}(i,m,l),\boldsymbol{x}(i',m',l'),\boldsymbol{s},\boldsymbol{s}',\boldsymbol{y})$.
3) for each $(\boldsymbol{v}(i),\boldsymbol{x}(i,m',l'))$ with $(m',l')\neq(m,l)$ such that $D(P_{VX'S'Y}||P_{VX'}\times P_{S'}\times W)\leq\eta$ for some $\boldsymbol{s}'\in\mathcal{S}^n$ satisfying $l^n(\boldsymbol{s}')\leq\Lambda$, we have $I(X,Y;X'|V,S)\leq\eta$. Here, (V,X,X',S,S',Y) are dummy random variables distributed according to the joint type of $(\boldsymbol{v}(i),\boldsymbol{x}(i,m,l),\boldsymbol{x}(i,m',l'),\boldsymbol{s},\boldsymbol{s}',\boldsymbol{y})$.

Remark 7: The decoder in the above coding scheme is almost the same as that in Definition 11 of [20], except that our decoder is used for a stochastic encoder. Similar to Section IV, by adopting a strict decoding rule, the proof in [20, Appendix I] can be directly applied here. In [20], the authors further define a common message decoder for the second receiver, which is not needed in this paper since in our model, the correctness of the eavesdropper's decoding of the common messages does not matter.

The coding scheme used in [20] is for an arbitrarily varying broadcast channel with a degraded message set. The main difference is that in our paper, the eavesdropper is not required to decode the common message while the legitimate receiver should decode the 'private message' correctly. In the single-user AVWC, one can regard the coding scheme as a superposition coding scheme, with some meaningless 'cloud center' codewords.

By the standard technique used in [20] and Fourier-Motzkin elimination, the legitimate receiver can decode the message correctly with high probability if

$$\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q|V)-\tau\leq\widetilde{R}\leq\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q|V)-\frac{2}{3}\tau$$

for some $\tau>0$ that can be arbitrarily small. Combining the constraint on R' in Lemma 8 and the constraint on $\widetilde{R}$, the reliable and secure communication can be achieved if

$$\begin{aligned}&\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q|V)-\max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(X;Z|S,V)-2\tau\\&\leq\frac{1}{n}\log|\mathcal{M}|=R\\&\leq\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q|V)-\max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(X;Z|S,V)-\frac{5}{3}\tau.\end{aligned}$$

Applying the prefixed channel argument and maximizing the achievable rate over all input distributions satisfying the constraints, the proof is completed. □

VI. Examples

In this section, we give some numerical examples to better illustrate our main results. In Subsection VI-A, we give numerical results of the stochastic-encoder code achievable rate and random code achievable rate of the AVWC with state constraint. In Subsection VI-B, we consider a special case of the AVWC, named severely less noisy AVWC, and use a numerical example to show that the stochastic-encoder code capacity can be strictly smaller than the random code capacity. In Subsection VI-C, it is proved that the superposition coding scheme used in this paper improves the secrecy achievable rate compared to the coding scheme without 'cloud center' in some cases.

A. Binary AVWC With State Constraint

Consider an AVWC whose main channel satisfies

$$Y=X+K\cdot S,\tag{37}$$

where $X\sim Bernoulli(p), K\sim Bernoulli(\theta), S\sim Bernoulli(q)$. The wiretap channel is a degraded version of the main channel. The transition probability from $\mathcal{Y}$ to $\mathcal{Z}$ forms an erasure channel with erasure probability α. The channel is subject to state constraint Λ and the state cost function is Hamming weight, i.e.

$$l(s)=\begin{cases}1, & s=1,\\ 0, & s=0.\end{cases}$$

The transmission rate $R=\max\{\alpha-h(\theta\Lambda),0\}$ can be achieved by both random code and stochastic-encoder code with state constraint Λ if $0\leq\theta\Lambda<\frac{1-\sqrt{1-\alpha}}{2}$ and $R=0$ otherwise.

The proof is given in Appendix F. In our previous work [28], a random code inner bound of such a model without state constraint is provided. The achievable rate in this example is almost the same as our previous result [28, Proposition 2], except that θ is replaced by $\theta\Lambda$ since in this case, the range of q is limited to $0\leq q\leq\Lambda$ due to state constraint. Based on this result, we have the following corollary.

Corollary 1: For an AVWC with state constraint Λ with a symmetrizable main channel, the stochastic-encoder secrecy capacity can still be positive.

In the next subsection, we use a numerical example to prove that the stochastic-encoder code capacity of the AVWC with constraints can be strictly smaller than the corresponding random code capacity.

B. Capacity of Severely Less Noisy AVWC With Constraints

The capacity results in Theorem 7 imply the following property of the arbitrarily varying wiretap channel with constraints, which is similar to the AVC with constraints.

Proposition 4: The stochastic-encoder code capacity of the AVWC with constraints can be positive but smaller than the corresponding random code capacity if the main channel of the AVWC is symmetrizable.

The proposition is proved by a numerical example, which is given in the rest of this subsection. In fact, we compare the

upper bound of the stochastic-encoder code capacity and the lower bound of the random code capacity, instead of computing the exact values of the capacities. The proof is completed by showing that the upper bound of the stochastic-encoder code capacity is strictly smaller than the lower bound of the random code capacity.

Example 2: Let $\mathcal{X} = \mathcal{S} = \mathcal{Y} = \{0,1\}, \mathcal{Z} = \{0,1,e\}$. The main channel $\mathcal{W} : \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$ is a two-state AVC $\mathcal{W} = \{W_1, W_2\}$ and the wiretap channel is a stationary erasure channel $E : \mathcal{X} \to \mathcal{Z}$, where

$$W_1 = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, W_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}, E = \begin{bmatrix} 1-e & e \\ e & 1-e \end{bmatrix}$$

and the cross probability $e = 0.45$. We further set $\Lambda = 0.8$ and $\Upsilon > \Lambda$. The constraint functions imposed on the input symbols and channel states are $\psi(x) = x$ and $l(s) = s$ for $x, s \in \{0,1\}$. It follows that the main channel is severely less noisy than the wiretap channel and the stochastic-encoder code capacity is strictly smaller than the random code capacity.

Proof: Note that the average channel can be expressed as

$$\begin{bmatrix} 1-\frac{q}{2} & \frac{q}{2} \\ \frac{1-q}{2} & \frac{1+q}{2} \end{bmatrix},$$

where $q = Pr\{S = 1\}$. We first prove that for any $q \in [0,1]$, the average channel W_q is less noisy than the wiretap channel E. This can be proved by an alternative definition of less noisy [31, Theorem 2]: A channel W with output random variable Y is less noisy than a channel E with output random variable Z if $I(X;Y) - I(X;Z)$ is a concave function of the input distribution P_X. Then we can verify by derivation that for any $q \in [0,1]$, $I(X;Y_q) - I(X;Z)$ is a concave function of the input distribution P_X. Hence, $\mathcal{W}$ is severely less noisy than E and the capacity results in Theorem 7 can be applied to this example. Note that the wiretap channel is a stationary channel, and the corresponding random code capacity is

$$\max_{P_X \in \mathcal{P}_\Upsilon(\mathcal{X})} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q) - I(X;Z)]. \tag{38}$$

Applying the channel model to formula (38) yields

$$C^{RC} = \max_{p \le \Upsilon} [\min_{q \le \Lambda} h(\frac{p+q}{2}) - (1-p)h(\frac{q}{2}) - ph(\frac{1-q}{2}) - (h((1-2e)p+e) - h(e))],$$

where $h(\cdot)$ is the entropy function and $p = P_X(1)$. Upon choosing input distribution $p = \frac{1}{2}$, the above formula reaches minimum at $q = \frac{1}{2}$. It follows that $h(\frac{1}{2}) - h(\frac{1}{4}) - (h(0.5) - h(0.45)) = 0.1815$ is achievable and hence $C^{RC} \ge 0.1815$.

Now let's move to the stochastic-encoder code capacity of this example. By Definition 6, the main channel $\mathcal{W}$ will only be symmetrized by identity matrix. Hence, $\Lambda_0(P_X) = p$. By the definition of less noisy and the technique used in Appendix E, it follows that

$$I(X;Y_q|V) - I_{q'}(X;Z|V,S) \le I(X;Y_q) - I_{q'}(X;Z|S)$$

for any $q, q' \in \mathcal{P}(\mathcal{S})$. Therefore, the stochastic-encoder code capacity satisfies

$$\begin{aligned} &\max_{\substack{P_{VX} \in \\ \mathcal{P}_{\Upsilon,\Lambda}(\mathcal{V},\mathcal{X})}} \min_{q,q' \in \mathcal{P}_\Lambda(\mathcal{S})} [I(X;Y_q|V) - I_{q'}(X;Z|V,S)] \\ &\le \max_{\substack{P_X \in \\ \mathcal{P}_{\Upsilon,\Lambda}(\mathcal{X})}} \min_{q,q' \in \mathcal{P}_\Lambda(\mathcal{S})} [I(X;Y_q) - I_{q'}(X;Z|S)] \\ &\le \max_{\substack{P_X \in \\ \mathcal{P}_{\Upsilon,\Lambda}(\mathcal{X})}} \min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} [I(X;Y_q)]. \end{aligned} \tag{39}$$

By derivation, the term $I(p,q) \triangleq I(X;Y_q) = h(\frac{p+q}{2}) - (1-p)h(\frac{q}{2}) - ph(\frac{1-q}{2})$ is concave in p and convex in q. Additionally, it follows that $p = 1/2$ and $q = 1/2$ is the saddle point of $I(p,q)$. For $\Upsilon > \Lambda = 0.8 > 0.5$, the right-hand side of the formula (39) reaches its maximum at $p = \Lambda$. So we have $C^{SC} \le I(\Lambda, 1/2) = 0.1228 < C^{RC}$. The proof is completed. □

With the help of Theorem 7 and Example 2, Proposition 4 shows that the stochastic-encoder code capacity of the AVWC with constraints can be strictly smaller than the corresponding random code capacity, while still being positive even when the main channel is symmetrizable, and this phenomenon is consistent with the original arbitrarily varying channel [8].

C. Comparison of the Achievable Rate Under Different Coding Schemes

In this section, we give an example that shows the advantage of using superposition coding over the AVWC with constraints.

Example 3 (Arbitrarily Varying Version of Proposition 1 in [25]): Let $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{S} = \mathcal{Y} = \{0,1\}$. The channel input is $X = (X_1, X_2)$ and U, V are the auxiliary random variables. The main channel to the legitimate receiver and the wiretap channel are expressed as

$$\begin{aligned} Y &= X_1 X_2 + N(1 - X_2) + S, \\ Z &= X_2, \end{aligned}$$

where $N \sim Bernoulli(\frac{1}{2})$. The channel is under input constraint $\Upsilon = 1/2$ and state constraint $\Lambda < 1/2$. The input cost and state cost functions are defined as $\psi(x) = \psi(x_1, x_2) = x_2$ and $l(s) = s$ for $x_1, x_2, s \in \{0,1\}$. For the AVWC with constraints defined as above, we have

$$\begin{aligned} &\max_{\substack{P_{VUX} \in \\ P_{\Upsilon,\Lambda}(\mathcal{V},\mathcal{U},\mathcal{X})}} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q|V) - \max_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I_q(U;Z|S,V)] \\ &\ge \frac{1}{2}(1 - h(\Lambda)) \\ &> \max_{\substack{P_{UX} \in \\ P_{\Upsilon,\Lambda}(\mathcal{U},\mathcal{X})}} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q) - \max_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I_q(U;Z|S)]. \end{aligned}$$

The analysis is similar to that in [25]. So we only give a proof outline in Appendix G. The proposition shows that the coding scheme used in Theorem 5 does improve the achievable rate for the AVWC with input constraint and state constraint.

VII. Conclusion

The strong secrecy problem of the AVWC with input and state constraints is investigated in this paper. Two lower bounds of the AVWC with constraints are established by applying the ordinary stochastic-encoder code and superposition stochastic-encoder code without the general assumption of the i.i.d. generated codebook. We further determine the random code secrecy capacity of the AVWC with state constraint.

For an ordinary AVC with constraints, the stochastic-encoder code capacity can be positive but strictly smaller than the corresponding random code capacity. In this paper, we find the property still holds for the AVWC with input and state constraints. To prove this, we determine the stochastic-encoder code and random code capacity for a special case of the AVWC, namely severely less noisy AVWC, and then use a numerical example to show the property.

Appendix A
Proof of Proposition 2

In this section, we prove the random code lower bound in Proposition 2. We adopt the random coding scheme in [24] with the help of the codeword properties (Lemma 3) in [8].

A. Error Analysis

1) Codebook Generation: Fix a type P_X on $\mathcal{X}$ such that $\mathbb{E}[X] = \sum_{x\in\mathcal{X}} P_X(x)\psi(x) \le \Upsilon$, where ψ is the cost function on $\mathcal{X}$ defined in Definition 2. Let $\boldsymbol{C} = \{\boldsymbol{X}(1), \ldots, \boldsymbol{X}(\widetilde{N})\}$ be a random codebook and $\{\mathcal{C}^\gamma\}_{\gamma\in\mathcal{I}}$ be the set of sample values of the random codebook $\boldsymbol{C}$ with $\mathcal{I}$ being the index set of the samples and μ being the distribution on $\mathcal{I}$. It follows that

$$Pr\{\boldsymbol{C} = \mathcal{C}^\gamma\} = \mu(\gamma) = \prod_{i=1}^{\widetilde{N}} \frac{1}{|T^n_{P_X}|},$$

where $\widetilde{N} = 2^{n\widetilde{R}}$ for some $\widetilde{R} > 0$.

2) Common Randomness: The common randomness randomly selects an index by distribution μ on index set $\mathcal{I}$. The codebook $\mathcal{C}^\gamma$ is then revealed to the sender and the receiver.

3) Encoding: The codebook $\mathcal{C}^\gamma$ is divided into $N = 2^{n(R-\tau)}$ parts $\{\mathcal{C}^\gamma(m)\}_{m=1}^N$ for some $R > 0$ and $\tau > 0$. To transmit message m, the encoder uniformly selects a codeword $\boldsymbol{x}(m,l)$ at random, which is the l^{th} codeword in the m^{th} sub-codebook $\mathcal{C}^\gamma(m)$.

4) Decoding: For any $\boldsymbol{s}$ with type q and $\delta > 0$, let $\mathcal{K}(\boldsymbol{s}) = \{(\boldsymbol{x},\boldsymbol{y}) : D(P_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{s}}||P_X \times q \times W) \le \delta\}$ and $\mathcal{K} = \cup_{\boldsymbol{s}\in\mathcal{S}^n(\Lambda)}\mathcal{K}(\boldsymbol{s})$, $\mathcal{K}_{m,l} = \{\boldsymbol{y} : (\boldsymbol{x}(m,l),\boldsymbol{y}) \in \mathcal{K}\}$. The decoding set is defined as

$$\mathcal{D}_{m,l} = \mathcal{K}_{m,l} \cap \Big(\bigcup_{(m',l')\neq(m,l)} \mathcal{K}_{m',l'}\Big)^c.$$

The decoder declares that message $\hat{m}$ is sent if there is a unique pair $(\hat{m},\hat{l})$ for codebook $\mathcal{C}^\gamma$ selected by common randomness such that $\boldsymbol{y} \in \mathcal{D}_{\hat{m},\hat{l}}$.

Given the partitioned codebook $\mathcal{C}^\gamma = \{\mathcal{C}^\gamma(m)\}_{m=1}^N$ and state sequence $\boldsymbol{s}$, the decoding error probability is bounded by

$$\begin{aligned} e(\boldsymbol{s}) &\le W^n(\mathcal{D}^c_{m,l}|\boldsymbol{x}(m,l),\boldsymbol{s}) \\ &\le W^n(\mathcal{K}^c_{m,l}|\boldsymbol{x}(m,l),\boldsymbol{s}) \\ &\quad + \sum_{(m',l')\neq(m,l)} W^n(\mathcal{K}_{m',l'}|\boldsymbol{x}(m,l),\boldsymbol{s}). \end{aligned} \tag{40}$$

Let $\mathcal{K}_{m,l}(\boldsymbol{s}) = \{\boldsymbol{y} : (\boldsymbol{x}(m,l),\boldsymbol{y}) \in \mathcal{K}(\boldsymbol{s})\}$. Applying the standard type argument [19],

$$\begin{aligned} &W^n(\mathcal{K}^c_{m,l}|\boldsymbol{x}(m,l),\boldsymbol{s}) \\ &\le W^n(\mathcal{K}^c_{m,l}(\boldsymbol{s})|\boldsymbol{x}(m,l),\boldsymbol{s}) \\ &= \sum_{\boldsymbol{y}:(\boldsymbol{x}(m,l),\boldsymbol{y})\in\mathcal{K}^c(\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}(m,l),\boldsymbol{s}) \\ &\overset{(a)}{\le} (n+1)^{|\mathcal{Y}|}\exp(-n(D(P_{XYS}||P_X\times q\times W) - I(X;S))) \\ &\overset{(b)}{\le} \exp(-n(\delta/2 - I(X;S))) \end{aligned} \tag{41}$$

for some $\delta > 0$, where (X,Y,S) are random variables distributed according to the joint type of $(\boldsymbol{x}(m,l),\boldsymbol{y},\boldsymbol{s})$ and (a) follows by the type counting lemma in [19], (b) follows by the definition of $\mathcal{K}(\boldsymbol{s})$. For each term in the sum in (40), it can be written as

$$W^n(\mathcal{K}_{m',l'}|\boldsymbol{x}(m,l),\boldsymbol{s}) = \sum_{\boldsymbol{y}:(\boldsymbol{x}(m',l'),\boldsymbol{y})\in\mathcal{K}} W^n(\boldsymbol{y}|\boldsymbol{x}(m,l),\boldsymbol{s}). \tag{42}$$

Denote the first term in (40) by $e_1(\boldsymbol{x}(m,l),\boldsymbol{s})$ and the term in the summation in (40) by $e_2(\boldsymbol{x}(m,l),\boldsymbol{x}(m',l'),\boldsymbol{s})$ for given (m',l'). Now considering the random codebook $\boldsymbol{C}$, the term $e_1(\boldsymbol{X}(m,l),\boldsymbol{s})$ is bounded by

$$\begin{aligned} &E[e_1(\boldsymbol{X}(m,l),\boldsymbol{s})] \\ &\le \exp(-n\delta/4) + Pr\{\boldsymbol{X}(m,l) \in \{\boldsymbol{x} : I(X;S) > \delta/4\}\} \\ &\overset{(a)}{\le} 2\exp(-n\delta/12), \end{aligned}$$

where (a) follows from the proof of Lemma 3 in [8]. For the term $e_2(\boldsymbol{X}(m,l),\boldsymbol{X}(m',l'),\boldsymbol{s})$, by independence we have

$$\begin{aligned} &E[e_2(\boldsymbol{X}(m,l),\boldsymbol{X}(m',l'),\boldsymbol{s})|\boldsymbol{X}(m,l) = \boldsymbol{x}(m,l)] \\ &= E[e_2(\boldsymbol{x}(m,l),\boldsymbol{X}(m',l'),\boldsymbol{s})|\boldsymbol{X}(m,l) = \boldsymbol{x}(m,l)] \\ &= \sum_{\boldsymbol{y}} W^n(\boldsymbol{y}|\boldsymbol{x}(m,l),\boldsymbol{s}) \left(\sum_{\boldsymbol{x}:(\boldsymbol{x},\boldsymbol{y})\in\mathcal{K}} Pr\{\boldsymbol{X}(m',l') = \boldsymbol{x}\}\right) \\ &\le 2^{-n(\inf_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q)-\epsilon)} \end{aligned}$$

for some $\epsilon > 0$. The average error probability of the random code is bounded by

$$\bar{\lambda}^{RC}(\mathcal{W},F,\Phi,\boldsymbol{s}) \le 2\cdot 2^{-n\delta/2} + \widetilde{N}\cdot 2^{-n(\inf_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q)-\epsilon)}.$$

Since $\widetilde{N} = 2^{n\widetilde{R}}$, the error probability is exponentially small if

$$\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q) - \tau \le \widetilde{R} \le \min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q) - \frac{2}{3}\tau$$

for some $\tau > 0$ such that $\frac{2}{3}\tau > \epsilon$.

B. Information Leakage

As proved in Section IV, one can construct a secure partition on a given good codebook $\mathcal{C}^\gamma$ if $R' > \max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X;Z|S)$. Now let $\boldsymbol{Z_s}^\gamma$ be the channel output with state sequence $\boldsymbol{s}\in\mathcal{S}^n(\Lambda)$ and codebook $\mathcal{C}^\gamma$ and $\mathcal{I}_g$ be the set of indices such that for any $\gamma\in\mathcal{I}_g$, $\mathcal{C}^\gamma$ is a good codebook. The information leakage is bounded by

$$\begin{aligned} &I(M,\boldsymbol{Z_s}|\Gamma)\\ &=\sum_\gamma I(M,\boldsymbol{Z_s}^\gamma|\Gamma=\gamma)\mu(\gamma)\\ &=\sum_{\gamma\in\mathcal{I}_g} I(M,\boldsymbol{Z_s}^\gamma|\Gamma=\gamma)\mu(\gamma)+\sum_{\gamma\notin\mathcal{I}_g} I(M,\boldsymbol{Z_s}^\gamma|\Gamma=\gamma)\mu(\gamma)\\ &\overset{(a)}{\le}\epsilon'+\zeta\cdot\log|M|\overset{(b)}{=}\varepsilon, \end{aligned}$$

where (a) follows by Lemma 5, (b) follows by Lemma 4 and the fact that ζ is a double exponentially small number. Combining the above subsections implies that the reliable and secure communication is achieved if

$$\begin{aligned} \min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q)-\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X;Z|S)-2\tau&\le\frac{1}{n}\log|\mathcal{M}|\\ &\le\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q)-\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X;Z|S)-\frac{5}{3}\tau. \end{aligned}$$

Maximizing over all possible input distribution and applying the prefixed channel argument discussed in Section IV-C, the proof is completed. □

APPENDIX B
PROOF OF LEMMA 2

In this section, we prove that a sequence randomly selected from a type set is typical under state sequences with high probability. Let $\boldsymbol{s}$ be a given sequence. For any given $s\in\mathcal{S}$ and random sequence $\boldsymbol{X}$ uniformly distributed over $T^n_{P_X}$, define an indicator random variable as follows.

$$V_{s,x}=\begin{cases}1, & |N(x,s|\boldsymbol{X},\boldsymbol{s})-N(s|\boldsymbol{s})P_X(x)|>n\delta,\\ 0, & \text{otherwise.}\end{cases} \tag{43}$$

Then we have

$$\begin{aligned} Pr\{V_{s,x}=1\}=Pr\{&N(x,s|\boldsymbol{X},\boldsymbol{s})-N(s|\boldsymbol{s})P_X(x)>n\delta\}\\ &+Pr\{N(x,s|\boldsymbol{X},\boldsymbol{s})-N(s|\boldsymbol{s})P_X(x)<-n\delta\}. \end{aligned} \tag{44}$$

By symmetry, it is sufficient to bound one of the above two formulas in the right-hand side of formula (44). It follows that

$$\begin{aligned} &Pr\{N(x,s|\boldsymbol{X},\boldsymbol{s})-N(s|\boldsymbol{s})P_X(x)>n\delta\}\\ &=\sum_{n_x=N(s|\boldsymbol{s})(P_X(x)+\delta')+1}^{N(s|\boldsymbol{s})} Pr\{N(x,s|\boldsymbol{X},\boldsymbol{s})=n_x\}\\ &=\sum_{n_x=N(s|\boldsymbol{s})(P_X(x)+\delta')+1}^{N(s|\boldsymbol{s})}\frac{\binom{N(s|\boldsymbol{s})}{n_x}\binom{n-N(s|\boldsymbol{s})}{nP_X(x)-n_x}}{\binom{n}{nP_X(x)}}, \end{aligned} \tag{45}$$

where $\delta'=n\delta/N(s|\boldsymbol{s})$. Note that if $P_X(x)=0$ or $P_X(x)+\delta'\ge 1$, the probability equals to 0 and there is nothing further to prove. Hence, we only need to consider the symbol pair (x,s) such that $P_X(x)+\delta'<1$. To bound the term $\binom{N(s|\boldsymbol{s})}{n_x}$, we need the following lemma.

Lemma 9: Let n be a positive real number and $p\in[0,1]$ such that np is an integer. $\binom{n}{np}$ is bounded by

$$(n+1)^{-2}2^{nH(p)}\le\binom{n}{np}\le 2^{nH(p)},$$

where $H(p)=-p\log p-(1-p)\log(1-p)$.

Proof: Consider the set of all n-length binary sequences with type $\hat{p}(1)=p$ and $\hat{p}(0)=1-p$. The set of sequences is denoted by $T^n_{\hat{p}}$. For each sequence $\boldsymbol{x}\in T^n_{\hat{p}}$, it follows that $N(1|\boldsymbol{x})=np$ and $N(0|\boldsymbol{x})=n(1-p)$. Now by Lemma 2.3 in [19],

$$(n+1)^{-2}2^{nH(p)}\le\binom{n}{np}=|T^n_{\hat{p}}|\le 2^{nH(p)}.$$

The proof is completed. □

Set $q=N(s|\boldsymbol{s})/n$ and $\lambda=n_x/n$, where n_x is introduced as in formula (45). It follows that

$$\delta<q<1-\delta, q\ P_X(x)+\delta<\lambda\le q, \lambda\le P_X(x)\le 1+\lambda-q. \tag{46}$$

By the above lemma, we bound the term

$$\binom{N(s|\boldsymbol{s})}{n_x}\le 2^{nqH(\frac{\lambda}{q})}$$

where the entropy function depends on $p(1)=\lambda/q$ and $p(0)=1-p(1)$. Similarly,

$$\binom{n-N(s|\boldsymbol{s})}{nP_X(x)-n_x}\le 2^{n(1-q)H(\frac{P_X(x)-\lambda}{1-q})},$$

where $\frac{P_X(x)-\lambda}{1-q}=(nP_X(x)-n_x)/(n-N(s|\boldsymbol{s}))$ and

$$\binom{n}{nP_X(x)}\ge(n+1)^{-2}2^{nH(P_X(x))}.$$

It follows that for any given $x\in\mathcal{X}$ and $s\in\mathcal{S}$,

$$\begin{aligned} &Pr\{N(x,s|\boldsymbol{X},\boldsymbol{s})-N(s|\boldsymbol{s})P_X(x)>N(s|\boldsymbol{s})\delta\}\\ &\le N(s|\boldsymbol{s})(n+1)^2\frac{2^{nqH(\frac{\lambda}{q})}2^{n(1-q))H(\frac{P_X(x)-\lambda}{1-q})}}{2^{nH(P_X)}}\\ &=N(s|\boldsymbol{s})(n+1)^2 2^{-n(H(P_X(x))-qH(\frac{\lambda}{q})-(1-q)H(\frac{P_X(x)-\lambda}{1-q}))}. \end{aligned} \tag{47}$$

For given $1>\delta>0$, define $f(x_1,x_2,x_3)=H(x_1)-x_2H(\frac{x_3}{x_2})-(1-x_2)H(\frac{x_1-x_3}{1-x_2})$ with $\delta\le x_2\le 1-\delta$, $x_1x_2+\delta\le x_3\le x_2$ and $x_3\le x_1\le 1+x_3-x_2$. Note that $\delta\le x_2\le 1-\delta$, $x_1x_2+\delta\le x_3\le x_2$ and $x_3\le x_1\le 1+x_3-x_2$ form a bounded closed region $\mathcal{R}$. By fact that $P_1\ne P_2$ implies $H(qP_1+(1-q)P_2)-qH(P_1)-(1-q)H(P_2)>0$, and here $P_1-P_2=\frac{x_3}{x_2}-\frac{x_1-x_3}{1-x_2}\ge\frac{\delta}{x_2(1-x_2)}\ge 4\delta>0$, we have

$$\nu\triangleq\min_{(x_1,x_2,x_3)\in\mathcal{R}} f(x_1,x_2,x_3)=f(x_1^*,x_2^*,x_3^*)>0$$

independent of n, where $(x_1^*,x_2^*,x_3^*)=\arg\min_{(x_1,x_2,x_3)\in\mathcal{R}} f(x_1,x_2,x_3)$. Now we return to

formula (47). Let $\mathcal{R}^*$ be the set of $(P_X(x), \lambda, q)$ satisfying constraints in formula (46). It follows that $\mathcal{R}^* \subseteq \mathcal{R}$, and then

$$\nu^* \triangleq \inf_{(P_X(x),q,\lambda)\in\mathcal{R}^*} f(P_X(x), q, \lambda) \geq \nu > 0$$

and formula (47) is upper bounded by $N(s|\boldsymbol{s})(n+1)^2 2^{-n\nu} < 2^{-n\nu'}$ for $\nu' > 0$. The above argument and the symmetry in formula (44) show that the probability that $V_{s,x} = 1$ for a give $x \in \mathcal{X}$ and $s \in \mathcal{S}$ is bounded by $Pr\{V_{s,x} = 1\} \leq 2 \cdot 2^{-n\nu'}$. Summing over all possible x and s, the probability that a randomly selected sequence from a type set $T^n_{P_X}$ is not typical under the given state sequence $\boldsymbol{s}$ satisfies

$$Pr\{\boldsymbol{X} \text{ is not typical under } \boldsymbol{s}\} \leq |\mathcal{X}||\mathcal{S}|2 \cdot 2^{-n\nu'}$$

for $\nu' > 0$ and sufficiently large n. The proof is completed. □

APPENDIX C
PARAMETERS VERIFICATION

In this section, we verify the parameters in formulas (26) and (33) that ensure the existence of the secure partitions in Lemma 5 and Lemma 8.

A. Verification of Parameters (26)

In this subsection, we verify that the parameters in (26) satisfy the conditions in Lemma 3 that ensure the existence of the secure partition.

1) Verification of $P(\mathcal{F}) \geq 1 - \epsilon$: For any $\boldsymbol{x} \in \mathcal{C} \cap T^n_{P_{X|Z,2\delta}}[\boldsymbol{z}, \boldsymbol{s}]$, $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and $\boldsymbol{z} \in \mathcal{B}(\boldsymbol{s})$, the conditional probability of codeword $\boldsymbol{x}$ satisfies

$$\begin{aligned}
&P^n_{\boldsymbol{X}|\boldsymbol{z},\boldsymbol{s}}(\boldsymbol{x}) \\
&= Pr\{\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{Z_s} = \boldsymbol{z}\} \\
&\overset{(a)}{\leq} \frac{V^n_{\boldsymbol{s}}(\boldsymbol{z}|\boldsymbol{x})}{2^{n\widetilde{R}}2^{-n\frac{\nu}{2}}\prod_{i=1}^{n} P_{Z_{s_i}}(z_i)} \\
&\overset{(b)}{\leq} \frac{2^{-n(H_q(Z|X,S)-\epsilon_1)}}{2^{n\widetilde{R}}2^{-n\frac{\nu}{2}}2^{-n(H_q(Z|S)+\epsilon_2)}} \\
&= 2^{-n(\widetilde{R}-I_q(X;Z|S)-\epsilon_1-\epsilon_2-\frac{\nu}{2})} \qquad (48) \\
&\leq 2^{-n(\widetilde{R}-\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X;Z|S)-\epsilon_1-\epsilon_2-\frac{\nu}{2})} \\
&\overset{(c)}{\leq} 2^{-n(R-\frac{\tau}{2})} = d^{-1}, \qquad (49)
\end{aligned}$$

where

- (a) follows from the fact that $|\mathcal{C}| = 2^{n\widetilde{R}}$ and the property of $\mathcal{B}_1(\boldsymbol{s})$;
- (b) follows by Lemma 1;
- (c) follows by the fact that $\widetilde{R} = R + R'$ and $R' > \max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X; Z|S)$.

Then we construct the subset $\mathcal{F}$ as $\mathcal{F} = \mathcal{C} \cap T^n_{P_{X|Z,2\delta}}[\boldsymbol{z}, \boldsymbol{s}]$ for $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and $\boldsymbol{z} \in \mathcal{B}(\boldsymbol{s})$. The properties of $\mathcal{B}_0$ in formula (23) imply that the condition $P(\mathcal{F}) \geq 1 - 2^{-n\nu/2}$ is satisfied.

For $P^n_{\boldsymbol{X}}$, note that all the codewords are uniformly distributed on the codebook, and $P^n_{\boldsymbol{X}}(\boldsymbol{x}) = 2^{-n\widehat{R}} < 2^{-n(R-\frac{\tau}{2})}$. Setting $\mathcal{F} = \mathcal{C}$, it follows that $P^n_{\boldsymbol{X}}(\mathcal{C}) = 1$.

2) Verification of $k \log k < \frac{\epsilon^2(1-\epsilon)d\log e}{2(1+\epsilon)\log 2|\mathcal{P}|}$: This condition ensures that the probability that the mapping exists is positive. To prove the desired result, we have

$$\begin{aligned}
&\frac{\epsilon^2(1-\epsilon)d\log e}{2(1+\epsilon)\log 2|\mathcal{P}|} \\
&= \frac{1}{2(1+\epsilon)\log 2|\mathcal{P}|} 2^{n(R-\nu-\frac{1}{2}\tau+\frac{\log((1-\epsilon)\log e)}{n})} \\
&\overset{(a)}{\geq} 2^{n(R-\nu-\frac{1}{2}\tau+\frac{\log((1-\epsilon)\log e)}{n}-\frac{\log(2(1+\epsilon)n\log 2|\mathcal{S}||\mathcal{Z}|)}{n})} \\
&\overset{(b)}{>} 2^{n(R-\frac{3}{4}\tau)} \overset{(c)}{>} k\log k,
\end{aligned}$$

where (a) follows because the size $|\mathcal{P}| \leq (|\mathcal{S}||\mathcal{Z}|)^n$, (b) and (c) are deduced by a sufficiently large n and a sufficiently small ν. The verification is completed. □

B. Verification of Parameters (33)

In this subsection, we verify the parameters in (33). Here we only prove that the condition $P(\mathcal{F}) \geq 1 - \epsilon$ holds and the verification of the other conditions are the same as that in the above subsection. For any $\boldsymbol{x} \in \mathcal{C}(i) \cap T^n_{X|VZ,3\delta}[\boldsymbol{v}(i), \boldsymbol{z}, \boldsymbol{s}]$, $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and $\boldsymbol{z} \in \mathcal{B}(\boldsymbol{s})$, the codeword $\boldsymbol{x}$ satisfies

$$\begin{aligned}
&P^n_{\boldsymbol{X}|\boldsymbol{v}(i),\boldsymbol{z},\boldsymbol{s}}(\boldsymbol{x}) \\
&= Pr\{\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{V} = \boldsymbol{v}(i), \boldsymbol{Z_s} = \boldsymbol{z}\} \\
&\leq \frac{P^n_{Z_{\boldsymbol{s}}|VX}(\boldsymbol{z}|\boldsymbol{v}(i), \boldsymbol{x})}{2^{n\widetilde{R}}2^{-n\frac{\nu}{2}}\prod_{i=1}^{n} P_{Z_{s_i}|V}(z_i|v_i)} \\
&\leq \frac{2^{-n(H_q(Z|V,X,S)-\epsilon_1)}}{2^{n\widetilde{R}}2^{-n\frac{\nu}{2}}2^{-n(H_q(Z|V,S)+\epsilon_2)}} \\
&= 2^{-n(\widetilde{R}-I_q(X;Z|V,S)-\epsilon_1-\epsilon_2-\frac{\nu}{2})} \\
&\leq 2^{-n(\widetilde{R}-\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X;Z|V,S)-\epsilon_1-\epsilon_2-\frac{\nu}{2})} \\
&\leq 2^{-n(R-\frac{\tau}{2})}.
\end{aligned}$$

Now we construct the subset $\mathcal{F}$ as $\mathcal{F} = \mathcal{C}(i) \cap T^n_{P_{X|VZ,2\delta}}[\boldsymbol{v}(i), \boldsymbol{z}, \boldsymbol{s}]$ for $\boldsymbol{v}(i) \in \mathcal{V}^n$, $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and $\boldsymbol{z} \in \mathcal{B}(\boldsymbol{s})$, the verification is completed. □

APPENDIX D
PROOF OF THEOREM 6 AND PROPOSITION 3

In this section, we prove the lower and upper bounds of the random code secrecy capacity of the AVWC with input constraint Υ and state constraint Λ.

A. Lower Bound

Similar to Proposition 1 and Proposition 2, we first set $U = X$ and prove the achievability and then apply the prefixed channel as we discussed in Section IV.

1) Codebook Generation: Fix input type $P_{VX}(v, x) = P_V(v)P_{V|X}(x|v)$ such that $\mathbb{E}[\psi(X)] = \sum_x \psi(x)\sum_v P_{VX}(v, x) \leq \Upsilon$. The random codebook for the common messages is denoted by $\boldsymbol{C}_C = \{\boldsymbol{V}(i)\}_{i=1}^{N_C}$, where $N_C = 2^{n\eta}$ for some $\eta > 0$ that can be arbitrarily small. For each i, the random codebook for the messages from sender is denoted by $\boldsymbol{C}(i) = \{\boldsymbol{X}(i, j)\}_{j=1}^{\widetilde{N}}$ where

$\boldsymbol{X}(i,j)$ is the j^{th} codeword in codebook $\boldsymbol{C}(i)$ such that $Pr\{\boldsymbol{X}(i,j)=\boldsymbol{x}|\boldsymbol{V}(i)=\boldsymbol{v}\}=1/|T^n_{P_{X|V}}[\boldsymbol{v}]|$, $\widetilde{N}=2^{n\widetilde{R}}$ for some $\widetilde{R}>0$. Let $\{\mathcal{C}_C^{\gamma_1}\}_{\gamma_1\in\mathcal{I}_1}$ and $\{\mathcal{C}^{\gamma_1,\gamma_2}(i)\}_{\gamma_1\in\mathcal{I}_1,\gamma_2\in\mathcal{I}_2}$ be collections of all the possible sample codebooks of random codebooks $\boldsymbol{C}_C$ and $\boldsymbol{C}(i)$ with $\mathcal{I}_j$ being the index set of the samples and μ_j being the distribution on $\mathcal{I}_j$ for $j=1,2$. It follows that

$$\begin{aligned}&Pr\{\boldsymbol{C}_C=\mathcal{C}_C^{\gamma_1}\}=\mu_1(\gamma_1)=\prod_{i=1}^{2^{n\eta}}\frac{1}{|T^n_{P_V}|},\\&Pr\{\boldsymbol{C}(i)=\mathcal{C}^{\gamma_1,\gamma_2}(i)|\boldsymbol{C}_C=\mathcal{C}_C^{\gamma_1}\}=\mu_2(\gamma_2)\\&=\prod_{j=1}^{2^{n\widetilde{R}}}\frac{1}{|T^n_{P_{X|V}}[\boldsymbol{v}(i)]|},\quad 1\le i\le N_C.\end{aligned}$$

2) Codebook Partition: For each codebook $\mathcal{C}^{\gamma_1,\gamma_2}(i)$ that is 'good' w.r.t. $\mathcal{C}_C^{\gamma_1}$ and real number $R'>\max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})}I_{q'}(X;Z|S,V)$, the codebook $\mathcal{C}^{\gamma_1,\gamma_2}(i)$ is divided into $2^{n(R-\tau)}$ sub-codebooks with $\tau>0$ by a partition satisfying Lemma 8. Codewords in each sub-codebook are indexed by a set $\mathcal{K}=[1,\ldots,2^{nR'}]$ such that $R'=\widetilde{R}-R$. Denote the sub-codebooks after partition by $\{\mathcal{C}^{\gamma_1,\gamma_2}(i,m)\}_{m=1}^{|\mathcal{M}|}$, where $\mathcal{M}$ is the message set.

3) Common Randomness and Encoding: The common randomness randomly selects indices γ_1 and γ_2 from index sets $\mathcal{I}_1$ and $\mathcal{I}_2$ by distributions μ_1 and μ_2, respectively. To transmit a message m, the sender selects a codeword $\boldsymbol{v}(i)$ uniformly at random from codebook $\mathcal{C}_C^{\gamma_1}$ and a codeword $\boldsymbol{x}(i,m,k)$ from subcodebook $\mathcal{C}^{\gamma_1,\gamma_2}(i,m)$, which is the k^{th} codeword in the m^{th} sub-codebook. The codebooks $(\mathcal{C}_C^{\gamma_1},\mathcal{C}^{\gamma_1,\gamma_2}(i))$ are revealed to the sender and receiver.

4) Decoding: For any $\boldsymbol{s}$ with type q and $\delta>0$, let $\mathcal{K}(\boldsymbol{s})=\{(\boldsymbol{v}(i),\boldsymbol{x}(i,m,k),\boldsymbol{y}):D(P_{\boldsymbol{v}(i)\boldsymbol{x}(i,m,k)\boldsymbol{y}\boldsymbol{s}}||P_{VX}\times q\times W)\le\delta\}$ and $\mathcal{K}=\cup_{\boldsymbol{s}\in\mathcal{S}^n(\Lambda)}\mathcal{K}(\boldsymbol{s}),\mathcal{K}_{i,m,k}=\{\boldsymbol{y}:(\boldsymbol{v}(i),\boldsymbol{x}(i,m,k),\boldsymbol{y})\in\mathcal{K}\}$. The decoding set is defined as

$$\mathcal{D}_{i,m,k}=\mathcal{K}_{i,m,k}\bigcap\left(\bigcup_{(i',m',k')\neq(i,m,k)}\mathcal{K}_{i',m',k'}\right)^c.$$

The decoder declares that message $\hat{m}$ is sent if there is a unique pair $(\hat{i},\hat{m},\hat{k})$ for codebook $(\mathcal{C}_C^{\gamma_1},\mathcal{C}^{\gamma_1,\gamma_2}(i))$ selected by common randomness such that $\boldsymbol{y}\in\mathcal{D}_{\hat{i},\hat{m},\hat{k}}$.

a) Error analysis: The constraint imposed on the input type implies that for each codeword $\boldsymbol{x}$,

$$\begin{aligned}\psi^n(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^n\psi(x_i)&=\sum_x\psi(x)P_X(x)\\&=\sum_x\psi(x)\sum_v P_{VX}(v,x)\\&=\sum_v P_V(v)\sum_x\psi(x)P_{X|V}(x|v)\\&=\mathbb{E}[\psi(X)]\le\Upsilon.\end{aligned}$$

The error analysis is similar to the proof of Proposition 2 in Appendix A and is omitted here. Applying Fourier-Motzkin elimination, it follows that the message can be decoded correctly if

$$\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q|V)-\tau\le\widetilde{R}\le\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q|V)-\frac{2}{3}\tau.$$

for some $\tau>0$ that can be arbitrarily small.

b) Information leakage: By Lemma 8, if $R'>\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I_q(X;Z|V,S)$, for any given state sequence $\boldsymbol{s}$ and codebook $\mathcal{C}^{\gamma_1,\gamma_2}(i)$ that is 'good' w.r.t. $\mathcal{C}_C$, there is a partition $g_i^{\gamma_1,\gamma_2}$ on codebook $\mathcal{C}^{\gamma_1,\gamma_2}(i)$ for $1\le i\le N_C$ such that the information leakage is bounded by $I(M;\boldsymbol{Z}_{\boldsymbol{s}}^{\gamma_1,\gamma_2})<\varepsilon$, where $\boldsymbol{Z}_{\boldsymbol{s}}^{\gamma_1,\gamma_2}$ is the channel output under state $\boldsymbol{s}$ with codebooks $(\mathcal{C}_C^{\gamma_1},\mathcal{C}^{\gamma_1,\gamma_2}(i))$ for some $1\le i\le N_C$. Let $\gamma\triangleq(\gamma_1,\gamma_2)$ and $\mathcal{I}_g$ be the set of index pair $\gamma=(\gamma_1,\gamma_2)$ such that $(\mathcal{C}_C^{\gamma_1},\mathcal{C}^{\gamma_1,\gamma_2}(i))$ are 'good superposition codebooks' for some $1\le i\le N_C$. The random code information leakage is upper bounded by

$$\begin{aligned}I(M;\boldsymbol{Z}_{\boldsymbol{s}}|\Gamma)&=\sum_{\gamma_1,\gamma_2}\mu_1(\gamma_1)\mu_2(\gamma_2)I(M;\boldsymbol{Z}_{\boldsymbol{s}}^{\gamma_1,\gamma_2}|\Gamma=\gamma)\\&\overset{(a)}{<}\varepsilon+\sum_{(\gamma_1,\gamma_2)\notin\mathcal{I}_g}\mu_1(\gamma_1)\mu_2(\gamma_2)\log|\mathcal{M}|\overset{(b)}{<}\varepsilon',\end{aligned}$$

where (a) follows by Lemma 8 and $I(M;\boldsymbol{Z}_{\boldsymbol{s}}^{\gamma})\le\log|\mathcal{M}|$ for any γ, (b) follows from the fact that $(\mathcal{C}_C^{\gamma_1},\mathcal{C}^{\gamma_1,\gamma_2}(i))$ are 'good superposition codebooks' with probability of at least $1-\zeta$, where ζ is a double exponentially small number. Hence, for give distribution P_{VX} defined in 'Codebook generation', the reliable and secure communication is achieved if

$$\begin{aligned}&\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q|V)-\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I_q(X;Z|V,S)-2\tau\\&\le\frac{1}{n}\log|\mathcal{M}|\\&\le\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I(X;Y_q|V)-\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})}I_q(X;Z|V,S)-\frac{5}{3}\tau.\end{aligned}$$

Maximizing over all possible input distribution and applying the standard prefixed channel argument discussed in Section IV-C, the proof of the achievability is completed. □

B. Upper Bound

To prove the converse part, we adopt the technique used in [32]. The following three auxiliary lemmas play an important role in the proof.

Lemma 10 (Lemma 4.5 in [32]): For every $\eta>0$, there exists a $\delta,0<\delta<\Lambda$, such that for all $q\in\mathcal{P}_\Lambda(\mathcal{S})$, there exists a $q'\in\mathcal{P}_{\Lambda-\delta}(\mathcal{S})$ with $d(q,q')=\sum_s|q(s)-q'(s)|\le\eta$.

Based on Lemma 10, one can further derive the following lemma.

Lemma 11: For every $\epsilon_1,\epsilon_2>0$, there exists a $0<\delta<\Lambda$ such that for any P_X,

$$\begin{aligned}&\min_q I(X;Y_q)-\max_{q'}I_{q'}(X;Z|S)\\&<\min_{q''}I(X;Y_{q''})-\max_{q'''}I_{q'''}(X;Z|S)+\epsilon,\end{aligned}$$

where $\epsilon=\epsilon_1+\epsilon_2$, q'' and q''' run over $\mathcal{P}_\Lambda(\mathcal{S})$ and q,q' run over $\mathcal{P}_{\Lambda-\delta}(\mathcal{S})$.

Proof: The proof of Lemma 4.6 in [32] implies that for given $\epsilon_1 > 0$ and $q''_* \in \mathcal{P}_\Lambda(\mathcal{S})$ such that $I(X;Y_{q''_*}) = \inf_{q''} I(X;Y_{q''})$, one can always find a $q \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})$ satisfying $d(q,q''_*) \le \eta$ and then $I(X;Y_q) < I(X;Y_{q''}) + \epsilon_1$. Similarly, let $q'''_* \in \mathcal{P}_\Lambda(\mathcal{S})$ such that $I_{q'''_*}(X;Z|S) = \max_{q'''} I_{q'''}(X;Z|S)$ and then there is a $q' \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})$ such that $d(q',q'''_*) \le \frac{\epsilon_2}{\log|\mathcal{X}|}$ for $\epsilon_2 > 0$. It follows that

$$\begin{aligned}
&|I_{q'''_*}(X;Z|S) - I_{q'}(X;Z|S)| \\
&\le \sum_s |q'''_*(s) I_{q'''_*}(X;Z_s) - q'(s) I_{q'}(X;Z_s)| \\
&\le \log|\mathcal{X}| \sum_s |q'''_*(s) - q'(s)| \le \epsilon_2.
\end{aligned}$$

Setting $\epsilon = \epsilon_1 + \epsilon_2$, the proof is completed. □

Now define the set of real numbers $\mathcal{R}_\Lambda \triangleq \{R : R \le \min_{q,q' \in \mathcal{P}_\Lambda(\mathcal{S})} \max_{\mathcal{P}(\mathcal{X})} [I(X;Y_q) - I_{q'}(X;Z|S)]\}$.

Lemma 12: For any $0 < \Lambda < l_{max}$, it follows that

$$\bigcap_{0<\delta<\Lambda} \mathcal{R}_{\Lambda-\delta} = \mathcal{R}_\Lambda.$$

The proof of Lemma 12 is the same as Lemma 4.4 in [32]. Let $\mathcal{R}_\Lambda^O$ be the set of real numbers satisfying

$$R \le \min_{q,q' \in \mathcal{P}_\Lambda(\mathcal{S})} \max_{P_{VUX} \in \mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X})} [I(U;Y_q|V) - I_{q'}(U;Z|V,S)].$$

By Lemma 12, it is sufficient to prove that the capacity falls into $\mathcal{R}_{\Lambda-\delta}^O$ for any $0 < \delta < \Lambda$. Let (F,Φ,Γ) be a random code over the AVWC with state constraint Λ. By Definition 5, the corresponding average decoding error probability and information leakage are bounded by

$$\bar{\lambda}^{RC}(\mathcal{W},F,\Phi,\boldsymbol{s}) < \varepsilon,\ I(M;\boldsymbol{Z_s}|\Gamma) < \varepsilon \tag{50}$$

for any $\boldsymbol{s} \in \mathcal{S}^n(\Lambda)$ and $\varepsilon > 0$.

To prove the converse, consider a channel $(\overline{W},\overline{E})$ with state sequence satisfying $Pr\{\boldsymbol{S} = \boldsymbol{s}\} = q^n(\boldsymbol{s}) = \prod_{i=1}^n q(s_i)$ and state constraint $E[l(S_i)] = \sum_s q(s) l(s) \le \Lambda - \delta$ for $1 \le i \le n$ and some $0 < \delta < \Lambda$. It follows that the random code (F,Φ,Γ) that achieves reliable and secure transmission can also be used over the channel $\overline{W}$. We first prove that the random code (F,Φ,Γ) can also achieve reliable transmission over the channel $\overline{W}$. For any distribution q^n, the average decoding error probability is

$$\begin{aligned}
&\mathbb{E}[\bar{\lambda}^{RC}(\overline{W},F,\Phi,\boldsymbol{S})] \\
&= \sum_{s^n} q^n(\boldsymbol{s}) \bar{\lambda}^{RC}(\overline{W},F,\Phi,\boldsymbol{s}) \\
&\le \sum_{s^n \in \mathcal{S}^n(\Lambda)} q^n(\boldsymbol{s}) \bar{\lambda}^{RC}(\overline{W},F,\Phi,\boldsymbol{s}) + Pr\{\boldsymbol{S} \notin \mathcal{S}^n(\Lambda)\} \\
&\le \varepsilon + Pr\{l^n(\boldsymbol{S}) \ge \Lambda\}.
\end{aligned}$$

Applying Hoeffding's inequality implies that

$$\begin{aligned}
&Pr\{l^n(\boldsymbol{S}) \ge \Lambda\} \\
&= Pr\{\frac{1}{n}\sum_{i=1}^n l(S_i) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[l(S_i)] \ge \Lambda - \mathbb{E}[l^n(\boldsymbol{S})]\} \\
&\le Pr\{\frac{1}{n}\sum_{i=1}^n l(S_i) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[l(S_i)] \ge \delta\} \\
&\le e^{-2\delta^2 n/l_{max}^2},
\end{aligned} \tag{51}$$

and then the error probability $\mathbb{E}[\bar{\lambda}^{RC}(\overline{W},F,\Phi,\boldsymbol{S})] \le \varepsilon + e^{-2\delta^2 n/l_{max}^2} < \varepsilon_1$ for some $\varepsilon_1 > 0$. Next we bound the average information leakage of the transmission over the channel $\overline{W}$ using the random code (F,Φ,Γ). Let $\widetilde{\boldsymbol{S}} = (\widetilde{S}_1,\widetilde{S}_2,\ldots,\widetilde{S}_n)$ be a random state sequence satisfying $Pr\{\widetilde{\boldsymbol{S}} = \widetilde{\boldsymbol{s}}\} = \prod_{i=1}^n q'(s_i)$. The average information leakage is bounded by

$$\begin{aligned}
I(M;\boldsymbol{Z}|\widetilde{\boldsymbol{S}},\Gamma) &= \sum_{\widetilde{\boldsymbol{s}} \in \mathcal{S}(\Lambda)} Pr\{\widetilde{\boldsymbol{S}} = \widetilde{\boldsymbol{s}}\} I(M;\boldsymbol{Z}|\widetilde{\boldsymbol{S}} = \widetilde{\boldsymbol{s}},\Gamma) \\
&\quad + \sum_{\widetilde{\boldsymbol{s}} \notin \mathcal{S}(\Lambda)} Pr\{\widetilde{\boldsymbol{S}} = \widetilde{\boldsymbol{s}}\} I(M;\boldsymbol{Z}|\widetilde{\boldsymbol{S}} = \widetilde{\boldsymbol{s}},\Gamma) \\
&\overset{(a)}{\le} \varepsilon + \log|\mathcal{M}| \cdot Pr\{l(\widetilde{\boldsymbol{S}}) > \Lambda\} \\
&\overset{(b)}{\le} \varepsilon + \log|\mathcal{M}| \cdot e^{-2n\delta^2/l_{max}^2} \le \varepsilon_2,
\end{aligned}$$

for some arbitrarily small $\varepsilon_2 > 0$, where (a) follows by formula (50), (b) follows by formula (51).

By the independence between the message M and the state sequence $\widetilde{\boldsymbol{S}}$, we also have $I(M;\boldsymbol{Z}|\widetilde{\boldsymbol{S}},\Gamma) = I(M;\widetilde{\boldsymbol{S}},\boldsymbol{Z}|\Gamma)$. Setting $\boldsymbol{Y}_q = (Y_{q,1},\ldots,Y_{q,n})$, by Fano's inequality, the secrecy capacity can be upper bounded by

$$\begin{aligned}
&nR \\
&= H(M) \\
&\le I(M;\boldsymbol{Y}_q|\Gamma) + \epsilon \\
&\le I(M;\boldsymbol{Y}_q|\Gamma) - I(M;\boldsymbol{Z}|\widetilde{\boldsymbol{S}},\Gamma) + \varepsilon_2 + \epsilon \\
&\overset{(a)}{=} \sum_{i=1}^n [I(M;Y_q^i,\widetilde{S}_{i+1}^n,Z_{i+1}^n|\Gamma) - I(M;Y_q^{i-1},\widetilde{S}_i^n,Z_i^n|\Gamma)] \\
&\quad + \epsilon + \varepsilon_2 \\
&= \sum_{i=1}^n [I(M;Y_{q,i}|Y_q^{i-1},\widetilde{S}_{i+1}^n,Z_{i+1}^n,\Gamma) \\
&\qquad - I(M;\widetilde{S}_i,Z_i|Y_q^{i-1},\widetilde{S}_{i+1}^n,Z_{i+1}^n,\Gamma)] + \epsilon + \varepsilon_2 \\
&\overset{(b)}{=} \sum_{i=1}^n [I(M;Y_{q,i}|\widetilde{V}_i) - I(M;\widetilde{S}_i,Z_i|\widetilde{V}_i)] + \epsilon + \varepsilon_2 \\
&\overset{(c)}{=} \sum_{i=1}^n [I(U_i;Y_{q,i}|\widetilde{V}_i) - I(U_i;\widetilde{S}_i,Z_i|\widetilde{V}_i)] + \epsilon + \varepsilon_2,
\end{aligned} \tag{52}$$

where (a) follows by [33, formulas (9) and (11)](also see formula (134) in [14]), $Y_q^i = (Y_{q,1},Y_{q,2},\ldots,Y_{q,i})$, $Z_{i+1}^n = (Z_{i+1},Z_{i+2},\ldots,Z_n)$, (b) follows by setting $\widetilde{V}_i = (Y_q^{i-1},\widetilde{S}_{i+1}^n,Z_{i+1}^n,\Gamma)$, and (c) follows by setting $U_i = (M,\widetilde{V}_i)$. For simplicity, we write $\widetilde{S}_i$ by S_i in the rest of this subsection, and formula (52) can be written as

$$R \le \frac{1}{n}\sum_{i=1}^n [I(U_i;Y_{q,i}|\widetilde{V}_i) - I_{q'}(U_i;S_i,Z_i|\widetilde{V}_i)] + \epsilon + \varepsilon_2 \tag{53}$$

$$\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^n [I(U_i;Y_{q,i}|\widetilde{V}_i,J=i) - I_{q'}(U_i;S_i,Z_i|\widetilde{V}_i,J=i)] \\
&\quad + \epsilon + \varepsilon_2 \\
&= [I(U_J;Y_{q,J}|\widetilde{V}_J,J) - I_{q'}(U_J;S_J,Z_J|\widetilde{V}_J,J)] + \epsilon + \varepsilon_2 \\
&\overset{(a)}{=} [I(U;Y_q|V) - I_{q'}(U;S,Z|V)] + \epsilon + \varepsilon_2,
\end{aligned} \tag{54}$$

where (a) follows by setting $U = U_J, V = (\widetilde{V}_J, J), S = S_J, Y_q = Y_{q,J}, Z = Z_J$. Here $V - U - (X, S) - (Y_q, Z)$ form a Markov chain and the conditional probabilities satisfy

$$\begin{aligned} &Pr\{Y_q = y|V = v, U = u, X = x\} \\ &= Pr\{Y_{q,J} = y|\widetilde{V}_J = \widetilde{v}, J = j, M = m, X_J = x\} \\ &\overset{(a)}{=} Pr\{Y_{q,j} = y|X_j = x\} = W_q(y|x) \end{aligned}$$

and

$$\begin{aligned} &Pr\{Z = z|V = v, U = u, X = x, S = s\} \\ &= Pr\{Z_J = z|\widetilde{V}_J = \widetilde{v}, J = j, M = m, X_J = x, S_J = s\} \\ &\overset{(b)}{=} Pr\{Z_j = z|X_j = x, S_j = s\} = E(z|x, s), \end{aligned}$$

where (a) and (b) follow by the Markov chain. Finally, maximizing formula (54) over all input distributions satisfying the input constraint yields

$$\begin{aligned} R &\le \max_{P_{VUX} \in \mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X})} [I(U; Y_q|V) - I_{q'}(U; S, Z|V)] + \varepsilon' \\ &= \max_{P_{VUX} \in \mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X})} [I(U; Y_q|V) - I_{q'}(U; Z|V, S)] + \varepsilon'. \end{aligned}$$

Since the distribution satisfying the state constraint can be arbitrary, we finally have

$$R \le \min_{q,q' \in \mathcal{P}_\Lambda(\mathcal{S})} \max_{\substack{P_{VUX} \in \\ \mathcal{P}_\Upsilon(\mathcal{V},\mathcal{U},\mathcal{X})}} [I(U; Y_q|V) - I_{q'}(U; Z|V, S)] + \varepsilon'.$$

The proof of the upper bound is completed. □

APPENDIX E
PROOF OF THEOREM 7

This section establishes the capacity results of the severely less noisy AVWC with input constraint Υ and state constraint Λ. Let (f, ϕ) be a stochastic-encoder code satisfying

$$\max_{\boldsymbol{s} \in \mathcal{S}^n(\Lambda)} \lambda^{SC}(\mathcal{W}, f, \phi, \boldsymbol{s}) < \varepsilon, \quad \max_{\boldsymbol{s} \in \mathcal{S}^n(\Lambda)} I(M; Z_{\boldsymbol{s}}) < \varepsilon.$$

The proof of achievability is straightforward by setting $U = X$ in Theorem 5. Thus, it is sufficient to prove the upper bound.

By formula (52), for any $0 < \delta < \Lambda$ and state distribution $q, q' \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})$ as defined in Appendix D-B, we have

$$\begin{aligned} R &\le \frac{1}{n} \sum_{i=1}^{n} [I(U_i; Y_{q,i}|\widetilde{V}_i) - I_{q'}(U_i; \widetilde{S}_i, Z_i|\widetilde{V}_i)] + \epsilon \\ &\overset{(a)}{=} I(U; Y_q|\widetilde{V}_J, J) - I_{q'}(U; Z|\widetilde{V}_J, S, J) + \epsilon \qquad (55) \\ &= I(U, X; Y_q|\widetilde{V}_J, J) - I_{q'}(U, X; Z|\widetilde{V}_J, S, J) \\ &\qquad - I(X; Y_q|U, \widetilde{V}_J, J) + I_{q'}(X; Z|U, \widetilde{V}_J, J, S) + \epsilon \\ &\overset{(b)}{\le} I(U, X; Y_q|\widetilde{V}_J, J) - I_{q'}(U, X; Z|\widetilde{V}_J, S, J) + \epsilon \\ &= I(U, X, \widetilde{V}_J; Y_q|J) - I_{q'}(U, X, \widetilde{V}_J; Z|S, J) - I(\widetilde{V}_J; Y_q|J) \\ &\qquad + I_{q'}(\widetilde{V}_J; Z|S, J) + \epsilon \\ &\overset{(c)}{\le} I(U, X, \widetilde{V}_J; Y_q|J) - I_{q'}(U, X, \widetilde{V}_J; Z|S, J) \\ &\overset{(d)}{=} I(X; Y_q|J) - I_{q'}(X; Z|S, J) + \epsilon, \end{aligned}$$

where (a) follows by setting $U = U_J, Y_q = Y_{q,J}, Z = Z_J, S = \widetilde{S} = S_J$, (b) and (c) follow by the definition of severely less noisy channel and $I_{q'}(U, X; Z|\widetilde{V}_J, S, J) \le \max_{s \in \mathcal{S}} I(U, X; Z|\widetilde{V}_J, J, S = s), I_{q'}(X; Z|S, J) \le \max_{s \in \mathcal{S}}(X; Z|J, S = s)$, (d) is due to the Markov chain $J - \widetilde{V}_J - U - X - (Y_q, Z)$. Since the distribution of S can be arbitrary, it follows that

$$R \le \min_{q \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})} I(X; Y_q|J) - \max_{q' \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})} I_{q'}(X; Z|S, J) + \epsilon. \qquad (56)$$

For the input constraint, we assume that $\widetilde{\Lambda}_0(P_{JX}) < \Lambda - \delta$. By the definition in formula (12), there exists a transition matrix $\widetilde{T}$ such that

$$\begin{aligned} \widetilde{\Lambda}_0(P_{JX}) &= \sum_{j,x,s} P_{JX}(j, x)\widetilde{T}(s|j, x)l(s) \\ &\overset{(a)}{=} \frac{1}{n} \sum_{j=1}^{n} P_{X|J}(x|j)\widetilde{T}(s|j, x)l(s) < \Lambda - \delta, \end{aligned}$$

where (a) follows by the fact that J is a random variable uniformly distributed on $[1 : n]$. It follows that for a random state sequence $\boldsymbol{S} = (S_1, S_2, \ldots, S_n)$ with conditional distribution $Pr\{\boldsymbol{S} = \boldsymbol{s}|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{J} = \boldsymbol{j}\} = \prod_{i=1}^{n} \widetilde{T}(s_i|x_i, j_i)$,

$$\begin{aligned} \mathbb{E}[l^n(\boldsymbol{S})] &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[l(S_i)] \\ &= \sum_{s \in \mathcal{S}} l(s) \frac{1}{n} \sum_{j=1}^{n} \sum_{x} \widetilde{T}(s|x, j)P_{X|J}(x|j) \\ &= \sum_{j,x,s} P_{JX}(j, x)\widetilde{T}(s|j, x)l(s) = \widetilde{\Lambda}_0(P_{JX}) < \Lambda - \delta. \end{aligned}$$

Now following the same argument in [8, Lemma 1], it can be proved that $\widetilde{\Lambda}_0(P_{JX}) < \Lambda - \delta$ leads to the average error probability $\bar{\lambda}^{SC}(\mathcal{W}, f, \phi, \boldsymbol{s}) > \frac{1}{4} - \delta'$ for some $\delta' > 0$. Generalizing random variable J and maximizing formula (56) over all possible P_{JX}, it follows that for any $0 < \delta < \Lambda$, the secrecy achievable rate is upper bounded by

$$\begin{aligned} R \le &\max_{\substack{P_{JX} \in \\ \mathcal{P}_{\Upsilon,\Lambda-\delta}(\mathcal{J},\mathcal{X})}} \min_{q \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})} I(X; Y_q|J) \\ &- \max_{q' \in \mathcal{P}_{\Lambda-\delta}(\mathcal{S})} I_{q'}(X; Z|S, J) + \epsilon. \qquad (57) \end{aligned}$$

The proof is completed by applying Lemma 10, Lemma 11 in Appendix D-B and Lemma 4.4 in [32], which is similar to the proof of Proposition 3. The proofs of the random code capacity and the stochastic-encoder code capacity with non-symmetrizable main channel follow similarly by setting $V = (\widetilde{V}_J, J)$ in formula (55). □

APPENDIX F
PROOF OF SECTION VI-A

A. Random Code Achievable Rate

The proof of the random code achievable rate is similar to our proof of Proposition 2 in [28]. Setting $U = X$ and

applying Proposition 1 along with the property of the erasure channel to the model, the objective function is

$$\max_{P_X}[\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(X;Y_q) - \max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})}(1-\alpha)I_{q'}(X;Y|S)].$$

We first deal with the term $I_{q'}(X;Y|S)$. For simplicity, set $q'(1) = q$ and $q'(0) = 1-q$. Differentiating $f(q) = I_{q'}(X;Y|S) = \sum_s q'(s)I(X;Y_s) = (1-q)h(p)+q(h(p*\theta)-h(\theta))$ with respect to q yields $f'(q) = h(p*\theta) - h(\theta) - h(p)$, where $h(p) = -p\log p - (1-p)\log(1-p)$ and $p*\theta = p(1-\theta)+(1-p)\theta$. It is shown that $f'(q) \le 0$ always holds no matter what value θ takes, so the function $f(q)$ decreases monotonically with respect to q and reaches its maximum at $q = 0$. By the definition of the cost function $l(\cdot)$, it follows that $l^n(\boldsymbol{s}) = q \le \Lambda$, which means that $q = 0$ is available under state constraint and $\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I_q(X;Y|S) = h(p)$.

To compute $I(X;Y_q)$, set $q(1) = q$ and rewrite it as $I(X;Y_q) = h(p*\theta q) - h(\theta q)$. This is a convex function over $0 \le \theta q \le 1$ and reaches its minimum 0 at $\theta q = 1/2$. When $\theta\Lambda \ge \frac{1}{2}$, there exists a q such that $\theta q = 1/2$, which leads to $R = 0$, and there is nothing further to prove. When $\theta\Lambda < \frac{1}{2}$, the minimum is achieved by $q = \Lambda$. Hence, our objective function is now

$$\max_p[h(p*\theta\Lambda) - h(\theta\Lambda) - (1-\alpha)h(p)]. \tag{58}$$

By differentiation, formula (58) is maximized at $p = 1/2$ when $0 \le \theta\lambda < \frac{1-\sqrt{1-\alpha}}{2}$ and $R = 0$ otherwise. This completes the proof of the random code achievable rate. □

B. Stochastic-Encoder Code Achievable Rate

To prove the stochastic achievable rate, we need to additionally consider whether the channel is symmetrizable. Note that Y, X, K, S are all binary random variables. Supposing there is a transition matrix $T : \mathcal{X} \to \mathcal{S}$ satisfying formula (3) and setting $x = 0$ and $x' = 1$, it follows that

$$\begin{aligned} &W(y|0,0)T(0|1) + W(y|0,1)T(1|1) \\ &= W(y|1,0)T(0|0) + W(y|1,1)T(1|0) \end{aligned} \tag{59}$$

for $y = 0, 1$. For simplicity, denote $T(0|0)$ by α_0 and $T(0|1)$ by α_1. Substituting $y = 0, 1$ and α_0, α_1 into (59) yields $\alpha_0 + \alpha_1 = 2 - 1/\theta$. Thus, the transition matrix T satisfying (3) exists, i.e. the channel is symmetrizable, if $\theta \ge 1/2$. In the following paragraphs, we consider the achievable rate in two cases, that is, whether the channel is symmetrizable.

Case 1: $\theta \ge \frac{1}{2}$.

In this case, the AVWC is symmetrizable. In addition, let $p \triangleq P_X(1)$. Recall that the existence of the stochastic-encoder code over the AVWC with state constraint depends on the value of $\Lambda_0^*(X) = \max_{P_X}\Lambda_0(P_X)$. It follows that

$$\begin{aligned} \Lambda_0^*(X) &= \max_{P_X}\min_T \sum_x\sum_s P_X(x)T(s|x)l(s) \\ &= \max_{P_X}\min_T [P_X(0)T(1|0) + P_X(1)T(1|1)] \\ &= \max_p \min_{\alpha_0,\alpha_1}[(1-p)(1-\alpha_0) + p(1-\alpha_1)] \\ &\overset{(a)}{=} \max_p \min_{\alpha_0}[(2p-1)\alpha_0 + \frac{p}{\theta} + 1 - 2p], \end{aligned}$$

where (a) follows by $\alpha_0+\alpha_1 = 2-1/\theta$. Note that $(2p-1)\alpha_0 + p/\theta + 1 - 2p$ is a linear function of α_0 and the minimum must be at $\alpha_0 = 0$ or $\alpha_0 = 1$, i.e. $\min_{\alpha_0}(2p-1)\alpha_0 + p/\theta + 1 - 2p = \min(p/\theta + 1 - 2p, p/\theta)$. The maximum over p is achieved when $p/\theta + 1 - 2p = p/\theta$ and $\Lambda_0^*(X) = 1/2\theta$ in this case. Thus, the stochastic-encoder code over the AVWC with state constraint can achieve a positive rate if $1/2\theta > \Lambda$, i.e. $\theta\Lambda < 1/2$. In the rest of this case, we assume that $\theta\Lambda < 1/2$, otherwise the stochastic-encoder code achievable rate is 0 and there is nothing further to prove. It follows that $\Lambda_0(p) = \min(p/\theta + 1 - 2p, p/\theta)$. In this case, our objective function is

$$\max_{p:\Lambda_0(p)\ge\Lambda}[h(p*\theta\Lambda) - h(\theta\Lambda) - (1-\alpha)h(p)], \tag{60}$$

which is almost the same as formula (58) except that the range of p is limited to $\theta\Lambda \le p \le (\Lambda-1)\theta/(1-2\theta)$. However, by the assumption that $\theta\Lambda < 1/2$, it follows that $\theta\Lambda < 1/2 < (\Lambda-1)\theta/(1-2\theta)$, which means that $p = 1/2$ is available and the stochastic-encoder code achievable rate in this case is the same as the random code achievable rate.

Case 2: $\theta < \frac{1}{2}$.

In this case, the AVWC is non-symmetrizable, and $\Lambda_0(p) = +\infty$ for any p. It is straightforward that the stochastic-encoder code achievable rate is the same as the random code achievable rate derived in this paper.

Combining Subsections A and B, we prove the random code and stochastic-encoder code achievable rates, respectively. □

Appendix G
Proof of Example 3

The proof is similar to that of Proposition 1 in [25], and we only give an outline here. To prove

$$\begin{aligned} &\max_{\substack{P_{VUX}\in \\ P_{\Upsilon,\Lambda}(\mathcal{V},\mathcal{U},\mathcal{X})}} [\min_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q|V) - \max_{q'\in\mathcal{P}_\Lambda(\mathcal{S})} I(U;Z|S,V)] \\ &\ge \frac{1}{2}(1-h(\Lambda)), \end{aligned} \tag{61}$$

first note that the wiretap channel is a stationary channel, and hence $\max_{q\in\mathcal{P}_\Lambda(\mathcal{S})} I(U;Z|S,V)] = I(U;Z|V)$. The main channel is symmetrized by the set of transition matrices $\mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{S}$ satisfying

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1-\alpha & \alpha \\ \frac{1}{2} & \frac{1}{2} \\ \alpha & 1-\alpha \end{bmatrix}$$

for any $0 \le \alpha \le 1$. Setting $V = X_2, U = X = (X_1, X_2)$ and $X_1 \sim Bernoulli(\frac{1}{2}), X_2 \sim Bernoulli(\frac{1}{2})$, by the definition of the cost function $l(s) = s$, we have $\widetilde{\Lambda}_0(P_{VUX}) = \frac{1}{2} > \Lambda$. It follows that

$$\begin{aligned} &\min_q I(U;Y_q|V) \\ &= \min_q I(X_1, X_2; Y_q|X_2) \\ &= \min_q [P_{X_2}(0)H(N+S) + P_{X_2}(1)H(X_1+S) \end{aligned}$$

$$
\begin{aligned}
&\quad -P_{X_2}(0)H(N+S) - P_{X_2}(1)H(S)] \\
&\overset{(a)}{=} \min_q [P_{X_2}(1)(h(p_1 * q) - h(q))] \\
&\overset{(b)}{=} \frac{1}{2}(1 - h(\Lambda)),
\end{aligned}
$$

where $h(p) = -p\log p - (1-p)\log(1-p), p * q = (1-p)*q + p*(1-q)$, (a) follows by setting $P_{X_1}(1) = p_1$, $q(1) = q$, (b) follows from the state constraint and the fact that $h(p * q) - h(q)$ is a convex function of q reaching minimum at $q = \frac{1}{2}$. Combining the fact that $I(U; Z|V) = 0$, formula (61) is proved.

To prove

$$
\begin{aligned}
&\frac{1}{2}(1-h(\Lambda)) \\
&\quad > \max_{\substack{P_{UX} \in \\ P_{\Upsilon,\Lambda}(\mathcal{U},\mathcal{X})}} [\min_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(U;Y_q) - \max_{q' \in \mathcal{P}_\Lambda(\mathcal{S})} I_{q'}(U;Z|S)],
\end{aligned} \tag{62}
$$

note that $\max_{q \in \mathcal{P}_\Lambda(\mathcal{S})} I(U;Z|S) = I(U;Z)$. It follows that

$$
\begin{aligned}
&\min_q [I(U;Y_q) - I(U;Z)] \\
&\overset{(a)}{=} \min_q [I(X_1;Y_q|X_2) - I(X_1;Y_q|X_2,U) - I(U;Z|Y_q)] \\
&\leq \min_q [I(X_1;Y_q|X_2)] \\
&\overset{(b)}{\leq} \frac{1}{2}(1 - h(\Lambda)),
\end{aligned}
$$

where the derivation of (a) is the same as that in [25], and the equality in (b) holds when the marginal distributions of X_1 and X_2 satisfy $P_{X_1}(1) = 1/2$ and $P_{X_2}(1) = 1/2$, $q(1) = \Lambda$ and $I(X_1;Y_q|X_2,U) = I(U;Z|Y_q) = 0$ for any $q \in \mathcal{P}(\mathcal{S})$. Furthermore, note that $Pr\{Y = 1|X_2 = 1\} = Pr\{X_1 + S = 1|X_2\} = 1/2$ and $Pr\{Y = 1|X_2 = 0\} = Pr\{N + S = 1|X_2 = 0\} = 1/2$ ($N \sim Bernoulli(\frac{1}{2})$), then we still have $X_2 \perp\!\!\!\perp Y$. Then following the same argument as that in [25, Proof of Proposition 1], the proof of the proposition is completed. □

REFERENCES

[1] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[2] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 44, no. 2, pp. 159–175, Jun. 1978.

[3] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 5, pp. 621–629, Sep. 1986.

[4] U. Pereg and Y. Steinberg, "The arbitrarily varying broadcast channel with causal side information at the encoder," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 757–779, Feb. 2020.

[5] U. Pereg and Y. Steinberg, "The arbitrarily varying degraded broadcast channel with causal side information at the encoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1033–1037.

[6] U. Pereg and Y. Steinberg, "The arbitrarily varying channel with colored Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3781–3817, Jun. 2021.

[7] I. Csiszar and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 1, pp. 27–34, Jan. 1988.

[8] I. Csiszar and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 2, pp. 181–193, Mar. 1988.

[9] U. Pereg and Y. Steinberg, "The arbitrarily varying channel under constraints with side information at the encoder," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 861–887, Feb. 2019.

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

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

[12] E. MolavianJazi, "Secure communications over arbitrarily varying wiretap channels," 2009. M.S. thesis, Dept. Elect. Eng., Graduate School, Univ. Notre Dame, Notre Dame, IN, USA, Dec. 2009.

[13] I. Bjelaković, H. Boche, and J. Sommerfeld, "Capacity results for arbitrarily varying wiretap channels," in *Information Theory, Combinatorics, and Search Theory*. Berlin, Germany: Springer, 2013, pp. 123–144.

[14] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Arbitrarily varying wiretap channels with type constrained states," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7216–7244, Dec. 2016.

[15] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.

[16] M. Wiese, J. Notzel, and H. Boche, "A channel under simultaneous jamming and eavesdropping attack—Correlated random coding capacities under strong secrecy criteria," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3844–3862, Jul. 2016.

[17] C. R. Janda, E. A. Jorswieck, M. Wiese, and H. Boche, "Arbitrarily varying wiretap channels with non-causal side information at the jammer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 938–943.

[18] J. Nötzel, M. Wiese, and H. Boche, "The arbitrarily varying wiretap channel—Secret randomness, stability, and super-activation," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3504–3531, Jun. 2016.

[19] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[20] E. Hof and S. I. Bross, "On the deterministic-code capacity of the two-user discrete memoryless arbitrarily varying general broadcast channel with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5023–5044, Nov. 2006.

[21] Y. Chen, D. He, C. Ying, and Y. Luo, "Strong secrecy of arbitrarily varying wiretap channels with constraints by stochastic code," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 843–848.

[22] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 1, pp. 42–48, Jan. 1985.

[23] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[24] J. A. Gubner and B. L. Hughes, "Nonconvexity of the capacity region of the multiple-access arbitrarily varying channel subject to constraints," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 3–13, Jan. 1995.

[25] S. Sreekumar, A. Bunin, Z. Goldfeld, H. H. Permuter, and S. Shamai, "The secrecy capacity of cost-constrained wiretap channels," *IEEE Trans. Inf. Theory*, vol. 67, no. 3, pp. 1433–1445, Mar. 2021.

[26] J. H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 2, pp. 212–226, Mar. 1981.

[27] D. He, Y. Luo, and W. Guo, "Arbitrarily varying wiretap channel: A new scheme for the proof of strong secrecy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 466–470.

[28] Y. Chen, D. He, and Y. Luo, "Strong secrecy of arbitrarily varying multiple access channels," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3662–3677, 2021.

[29] G. Kramer, *Topics in Multi-User Information Theory*. Delft, The Netherlands: Now, 2008.

[30] D. He and W. Guo, "Strong secrecy capacity of a class of wiretap networks," *Entropy*, vol. 18, no. 7, p. 238, Jun. 2016.

[31] M. van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 712–714, Mar. 1997.

[32] J. A. Gubner, "Deterministic codes for arbitrarily varying multiple-access channels," Ph.D. dissertation, Syst. Res. Center, Univ. Maryland, College Park, MD, USA, 1988.

[33] G. Kramer, "Teaching IT: An identity for the Gelfand-Pinsker converse," *IEEE Inf. Theory Soc. Newslett.*, vol. 61, no. 4, pp. 4–6, Dec. 2011.

Yiqi Chen received the B.E. degree from the Department of Communication Engineering, Xidian University, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. His current research interests include information theory and information theoretic security.

Dan He received the B.S. degree from East China Normal University in 2010, the M.S. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, in 2013, and the Ph.D. degree in information and communication engineering from Xidian University in 2017. He is currently a Senior Engineer with Intel Asia-Pacific Research and Development Ltd. His research interests include Shannon information theory, cloud computing, and edge computing.

Chenhao Ying received the B.E. degree from the Department of Communication Engineering, Xidian University, China, in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. His current research interests include mobile crowd sensing, communication coding algorithms, and wireless communications.

Yuan Luo (Member, IEEE) received the B.S. degree in applied mathematics and the M.S. and Ph.D. degrees in probability statistics from Nankai University, Tianjin, China, in 1993, 1996, and 1999, respectively. From July 1999 to April 2001, he held a post-doctoral position with the Institute of Systems Science, Chinese Academy of Sciences, Beijing, China. From May 2001 to April 2003, he held another post-doctoral position with the Institute for Experimental Mathematics, University of Duisburg-Essen, Essen, Germany. Since June 2003, he has been with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. Since 2006, he has been a Full Professor and the Vice Dean of the department (from 2016 to 2018 and since 2021). His current research interests include coding theory, information theory, and big data analysis.