Capacity Results
for Arbitrarily Varying Wiretap Channels

Igor Bjelaković, Holger Boche, and Jochen Sommerfeld

Lehrstuhl für theoretische Informationstechnik,
Technische Universität München, 80290 München, Germany
{igor.bjelakovic,boche,jochen.sommerfeld}@tum.de

Dedicated to the memory of Rudolf Ahlswede

Abstract. In this work the arbitrarily varying wiretap channel AVWC is studied. We derive a lower bound on the random code secrecy capacity for the average error criterion and the strong secrecy criterion in the case of a best channel to the eavesdropper by using Ahlswede's robustification technique for ordinary AVCs. We show that in the case of a non-symmetrisable channel to the legitimate receiver the deterministic code secrecy capacity equals the random code secrecy capacity, a result similar to Ahlswede's dichotomy result for ordinary AVCs. Using this we can derive that the lower bound is also valid for the deterministic code capacity of the AVWC. The proof of the dichotomy result is based on the elimination technique introduced by Ahlswede for ordinary AVCs. We further prove upper bounds on the deterministic code secrecy capacity in the general case, which results in a multi-letter expression for the secrecy capacity in the case of a best channel to the eavesdropper. Using techniques of Ahlswede, developed to guarantee the validity of a reliability criterion, the main contribution of this work is to integrate the strong secrecy criterion into these techniques.

Keywords: arbitrarily varying wiretap channels, strong secrecy, classes of attacks, jamming, active wiretapper.

1 Introduction

Models of communication systems taking into account both the requirement of security against a potential eavesdropper and reliable information transmission to legitimate receivers which suffer from channel uncertainty, have received much interest in current research. One of the simplest communication models with channel uncertainty are compound channels, where the channel realisation remains fixed during the whole transmission of a codeword. Compound wiretap channels were the topic of previous work of the authors [7], [8] and for example of [13], [9]. In the model of an arbitrarily varying wiretap channel AVWC the channel state to both the legitimate receiver and the eavesdropper varies from symbol to symbol in an unknown and arbitrary manner. Thus apart from eavesdropping the model takes into account an active adversarial jamming situation in

which the jammer chooses the states at her/his will. Then reliable transmission to the legitimate receiver must be guaranteed in the presence of the jammer.

In this paper we consider families of pairs of channels $\mathfrak{W} = \{(W_{s^n}, V_{s^n}) : s^n \in S^n\}$ with common input alphabets and possibly different output alphabets, where $s^n \in S^n$ denotes the state sequence during the transmission of a codeword. The legitimate users are connected via W_{s^n} and the eavesdropper observes the output of V_{s^n}. In our communication scenario the legitimate users have no channel state information. We derive capacity results for the AVWC $\mathfrak{W}$ under the average error probability criterion and a strong secrecy criterion. The investigation of the corresponding problem concerning the maximum error criterion is left as a subject of future investigations. However, we should emphasize that there is no full capacity result for an ordinary (i.e. without secrecy constraints) AVC for the maximum error criterion. Together with Wolfowitz in [6] Ahlswede determined the capacity for AVCs with binary output alphabets under this criterion. In [1] he showed that the general solution is connected to Shannon's zero error capacity problem [15].

Two fundamental techniques, called *elimination* and *robustification technique* discovered by Ahlswede will play a crucial role in this paper. In [2] he developed the *elimination technique* to derive the deterministic code capacity for AVCs under the average error probability criterion, which is either zero or equals its random code capacity, a result, which is called Ahlswede's dichotomy for single user AVCs. With the so-called *robustification technique* [3] in turn he could link random codes for the AVC to deterministic codes for compound channels. Further in the papers [4], [5] on *common randomness* in information theory Ahlswede together with Csiszár studied, inter alia, problems of information theoretic security by considering a model which enables secret sharing of a random key, in particular in the presence of a wiretapper. Because the arbitrarily varying wiretap channel AVWC combines both the wiretap channel and the AVC it is not surprising that we can use the aforementioned techniques to derive capacity results for the AVWC. The actual challenge of our work was to integrate the strong secrecy criterion in both the *elimination* and the *robustification technique*, approaches, both were developed to guarantee a reliability criterion. As it was shown in [8], compared with weaker secrecy criteria, the strong secrecy criterion ensures that the average error probability of every decoding strategy of the eavesdropper in the limit tends to one.

In Section 3.2 we give a lower bound on the random code secrecy capacity in the special case of a "best" channel to the eavesdropper. The proof is based on the *robustification technique* by Ahlswede [3] combined with results for compound wiretap channels given by the authors in [8].

In Section 3.3 we use the *elimination technique* [2], which is composed of the *random code reduction* and the *elimination of randomness* [10], to show that, provided that the channel to the legitimate receiver is non-symmetrisable, the deterministic code secrecy capacity equals the random code secrecy capacity and to give a condition when it is greater than zero. Thus we establish a result for the AVWC that is similar to that of Ahlswede's dichotomy result for ordinary AVCs.

As a consequence the above-mentioned lower bound on the random code secrecy capacity can be achieved by a deterministic code under the same assumptions.

In Section 3.4 we give a single-letter upper bound on the deterministic code secrecy capacity, which corresponds to the upper bound of the secrecy capacity of a compound wiretap channel. Moreover, by establishing an multi-letter upper bound on the secrecy capacity we can conclude to a multi-letter expression of the secrecy capacity of the AVWC in the special case of a best channel to the eavesdropper.

The lower bound on the secrecy capacity as well as other results were given earlier in [14] for a weaker secrecy criterion, but the proof techniques for the stronger secrecy criterion differ significantly, especially in the achievability part for the random codes.

2 Arbitrarily Varying Wiretap Channels

2.1 Definitions

Let A, B, C be finite sets and consider a non-necessarily finite family of channels $W_s : A \to \mathcal{P}(B)^1$, where $s \in S$ denotes the state of the channel. Now, given $s^n = (s_1, s_2, \ldots, s_n) \in S^n$ we define the stochastic matrix

$$W^n(y^n|x^n, s^n) := \prod_{i=1}^{n} W(y_i|x_i, s_i) := \prod_{i=1}^{n} W_{s_i}(y_i|x_i) \qquad (1)$$

for all $y^n = (y_1, \ldots, y_n) \in B^n$ and $x^n = (x_1, \ldots, x_n) \in A^n$. An arbitrarily varying channel is then defined as the sequence $\{\mathcal{W}^n\}_{n=1}^{\infty}$ of the family of channels $\mathcal{W}^n = \{W^n(\cdot|\cdot, s^n) : s^n \in S^n\}$. Now let $\mathcal{W}^n$ represent the communication link to a legitimate receiver to which the transmitter wants to send a private message, such that a possible second receiver should be kept as ignorant of that message as possible. We call this receiver the eavesdropper, which observes the output of a second family of channels $\mathcal{V}^n = \{V^n(\cdot|\cdot, s^n) : s^n \in S^n\}$ with an analogue definition of $V^n(\cdot|\cdot, s^n)$ as in (1) for $V_s : A \to \mathcal{P}(C), s \in S$. Then we denote the set of the two families of channels with common input by $\mathfrak{W} = \{(W_{s^n}, V_{s^n}) : s^n \in S^n\}$ and call it the arbitrarily varying wiretap channel. In addition, we assume that the state sequence s^n is unknown to the legitimate receiver, whereas the eavesdropper always knows which channel is in use.

A (n, J_n) code $\mathcal{C}_n$ for the arbitrarily varying wiretap channel $\mathfrak{W}$ consists of a stochastic encoder $E : \mathcal{J}_n \to \mathcal{P}(A^n)$ (a stochastic matrix) with a message set $\mathcal{J}_n := \{1, \ldots, J_n\}$ and a collection of mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. The average error probability of a code $\mathcal{C}_n$ is given by

$$e(\mathcal{C}_n) := \max_{s^n \in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j) W_{s^n}^n(D_j^c|x^n) \ . \qquad (2)$$

---

^{1} $\mathcal{P}(B)$ denotes the set of probability distributions on B.

A *correlated random* (n, J_n, Γ, μ) *code* $\mathcal{C}_n^{\mathrm{ran}}$ for the arbitrarily varying wiretap channel is given by a family of wiretap codes $\{\mathcal{C}_n(\gamma)\}_{\gamma \in \Gamma}$ together with a random experiment choosing γ according to a distribution μ on Γ. The mean average error probability of a random $(n.J_n, \Gamma, \mu)$ code $\mathcal{C}_n^{\mathrm{ran}}$ is defined analogously to the ordinary one but with respect to the random experiment choosing γ by

$$\bar{e}(\mathcal{C}_n^{\mathrm{ran}}) := \max_{s^n \in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{\gamma \in \Gamma} \sum_{x^n \in A^n} E^\gamma(x^n|j) W_{s^n}^n((D_j^\gamma)^c|x^n)\mu(\gamma) \ .$$

Definition 1. *A non-negative number R_S is an achievable secrecy rate for the AVWC $\mathfrak{W}$, if there is a sequence $(\mathcal{C}_n)_{n \in \mathbb{N}}$ of (n, J_n) codes such that*

$$\lim_{n \to \infty} e(\mathcal{C}_n) = 0 \ ,$$

$$\liminf_{n \to \infty} \frac{1}{n} \log J_n \geq R_S \ ,$$

and

$$\lim_{n \to \infty} \max_{s^n \in S^n} I(p_J; V_{s^n}^n) = 0 \ , \tag{3}$$

where J is a uniformly distributed random variable taking values in $\mathcal{J}_n$ and $I(p_J; V_{s^n}^n)$ is the mutual information of J and the output variable Z^n of the eavesdropper's channel $V_{s^n}^n$. The secrecy capacity then is given as the supremum of all achievable secrecy rates R_S and is denoted by $C_S(\mathfrak{W})$.

Analogously we define the secrecy rates and the secrecy capacity for random codes $C_{S,\mathrm{ran}}(\mathfrak{W})$, if we replace $\mathcal{C}_n$ by $\mathcal{C}_n^{\mathrm{ran}}$ in the above definition.

Definition 2. *A non-negative number R_S is an achievable secrecy rate for correlated random codes for the AVWC $\mathfrak{W}$, if there is a sequence $(\mathcal{C}_n^{\mathrm{ran}})_{n \in \mathbb{N}}$ of (n, J_n, Γ, μ) codes such that*

$$\lim_{n \to \infty} \bar{e}(\mathcal{C}_n^{\mathrm{ran}}) = 0 \ ,$$

$$\liminf_{n \to \infty} \frac{1}{n} \log J_n \geq R_S \ ,$$

and

$$\lim_{n \to \infty} \max_{s^n \in S^n} \sum_{\gamma \in \Gamma} I(p_J, V_{s^n}^n; \mathcal{C}(\gamma))\mu(\gamma) = 0 \ , \tag{4}$$

where $I(p_J, V_{s^n}^n; \mathcal{C}(\gamma))$ is the mutual information according to the code $\mathcal{C}(\gamma)$, $\gamma \in \Gamma$ chosen according to the distribution μ. The secrecy capacity then is given as the supremum of all achievable secrecy rates R_S and is denoted by $C_{S,\mathrm{ran}}(\mathfrak{W})$.

3   Capacity Results

3.1   Preliminaries

In what follows we use the notation as well as some properties of *typical* and *conditionally typical* sequences from [10]. For $p \in \mathcal{P}(A)$, $W : A \rightarrow \mathcal{P}(B)$, $x^n \in A^n$, and $\delta > 0$ we denote by $\mathcal{T}_{p,\delta}^n$ the set of typical sequences and by $\mathcal{T}_{W,\delta}^n(x^n)$ the set of conditionally typical sequences given x^n in the sense of [10].
   The basic properties of these sets that are needed in the sequel are summarised in the following three lemmata.

Lemma 1. *Fixing $\delta > 0$, for every $p \in \mathcal{P}(A)$ and $W : A \rightarrow \mathcal{P}(B)$ we have*

$$p^{\otimes n}(\mathcal{T}_{p,\delta}^n) \geq 1 - (n+1)^{|A|} 2^{-nc\delta^2}$$
$$W^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)|x^n) \geq 1 - (n+1)^{|A||B|} 2^{-nc\delta^2}$$

for all $x^n \in A^n$ with $c = 1/(2\ln 2)$. In particular, there is $n_0 \in \mathbb{N}$ such that for each $\delta > 0$ and $p \in \mathcal{P}(A)$, $W : A \rightarrow \mathcal{P}(B)$

$$p^{\otimes n}(\mathcal{T}_{p,\delta}^n) \geq 1 - 2^{-nc'\delta^2}$$
$$W^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)|x^n) \geq 1 - 2^{-nc'\delta^2}$$

holds with $c' = \frac{c}{2}$.

Proof. Standard Bernstein-Sanov trick using the properties of types from [10] and Pinsker's inequality. The details can be found in [16] and references therein for example. □

Recall that for $p \in \mathcal{P}(A)$ and $W : A \rightarrow \mathcal{P}(B)$, $pW \in \mathcal{P}(B)$ denotes the output distribution generated by p and W and that $x^n \in \mathcal{T}_{p,\delta}^n$ and $y^n \in \mathcal{T}_{W,\delta}^n(x^n)$ imply that $y^n \in \mathcal{T}_{pW,2|A|\delta}^n$.

Lemma 2. *Let $x^n \in \mathcal{T}_{p,\delta}^n$, then for $V : A \rightarrow \mathcal{P}(C)$*

$$|\mathcal{T}_{pV,2|A|\delta}^n| \leq \alpha^{-1}$$
$$V^n(z^n|x^n) \leq \beta \quad \text{for all} \quad z^n \in \mathcal{T}_{V,\delta}^n(x^n)$$

hold, where

$$\alpha = 2^{-n(H(pV)+f_1(\delta))} \tag{5}$$
$$\beta = 2^{-n(H(V|p)-f_2(\delta))} \tag{6}$$

with universal $f_1(\delta), f_2(\delta) > 0$ satisfying $\lim_{\delta \to \infty} f_1(\delta) = 0 = \lim_{\delta \to \infty} f_2(\delta)$.

Proof. Cf. [10].

The next lemma is a standard result from large deviation theory.

Lemma 3. *(Chernoff bounds) Let $Z_1, \ldots, Z_L$ be i.i.d. random variables with values in $[0, 1]$ and expectation $\mathbb{E}Z_i = \mu$, and $0 < \epsilon < \frac{1}{2}$. Then it follows that*

$$Pr\left\{ \frac{1}{L} \sum_{i=1}^{L} Z_i \notin [(1 \pm \epsilon)\mu] \right\} \leq 2 \exp\left(-L \cdot \frac{\epsilon^2 \mu}{3} \right),$$

where $[(1 \pm \epsilon)\mu]$ denotes the interval $[(1 - \epsilon)\mu, (1 + \epsilon)\mu]$.

For the optimal random coding strategy of the AVWC we need the *robustification technique* by Ahlswede [3] which is formulated as a further lemma. Therefore let Σ_n be the group of permutations acting on $(1, 2, \ldots, n)$. Then every permutation $\sigma \in \Sigma_n$ induces a bijection $\pi \in \Pi_n$ defined by $\pi : \mathcal{S}^n \to \mathcal{S}^n$ with $\pi(s^n) = (s_{\sigma(1)}, \ldots, s_{\sigma(n)})$ for all $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$ and Π_n denotes the group of these bijections.

Lemma 4. *(Robustification technique) If a function $f : \mathcal{S}^n \to [0, 1]$ satisfies*

$$\sum_{s^n \in \mathcal{S}^n} f(s^n) q(s_1) \cdot \ldots \cdot q(s_n) \geq 1 - \gamma \tag{7}$$

for all $q \in \mathcal{P}_0(n, \mathcal{S})$ and some $\gamma \in [0, 1]$, then

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi(s^n)) \geq 1 - 3 \cdot (n+1)^{|\mathcal{S}|} \cdot \gamma \quad \forall s^n \in \mathcal{S}^n . \tag{8}$$

Proof. The proof is given in [3].

To reduce the random code for the AVWC $\mathfrak{W}$ to a deterministic code we need the concept of symmetrisability, which was established for ordinary AVCs in the following representation by [12], [11].

Definition 3. *[11] An AVC is symmetrisable if for some channel $U : A \to S$*

$$\sum_{s \in S} W(y|x, s) U(s|x') = \sum_{s \in S} W(y|x', s) U(s|x) \tag{9}$$

for all $x, x' \in A$, $y \in B$.

A new channel defined by (9) then would be symmetric with respect to all $x, x' \in A$. The authors of [11] proved the following theorem which is a concretion of Ahlswede's dichotomy result for single-user AVC, which states that the deterministic code capacity C is either $C = 0$ or equals the random code capacity.

Theorem 1. *[11] $C > 0$ if and only if the AVC is non-symmetrisable. If $C > 0$, then*

$$C = \max_{p \in \mathcal{P}(A)} \min_{W \in \overline{\mathcal{W}}} I(p, W) \tag{10}$$

Here the RHS gives the random code capacity and $\overline{\mathcal{W}}$ denotes the convex closure of all channels W_s with $s \in S$, S finite or countable.

3.2   Random Code Construction

First let us define the convex hull of the set of channels $\{W_s : s \in S\}$ by the set of channels $\{W_q : q \in \mathcal{P}(S)\}$, where W_q is defined by

$$W_q(y|x) = \sum_{s \in S} W(y|x,s)q(s), \tag{11}$$

for all possible distributions $q \in \mathcal{P}(S)$. Accordingly we define V_q and its convex hull $\{V_q : q \in \mathcal{P}(S)\}$. Then we denote the convex closure of the set of channels $\{(W_s, V_s) : s \in S\}$ by $\overline{\mathfrak{W}} := \{(W_q, V_q) : q \in \mathcal{P}(\tilde{S}), \tilde{S} \subseteq S, \tilde{S} \text{ is finite}\}$. Occasionally, we restrict q to be from the set of all types $\mathcal{P}_0(n, S)$ of state sequences $s^n \in S^n$.

Lemma 5. *The secrecy capacity $C_S(\mathfrak{W})$ of the arbitrarily varying wiretap channel $AVWC\,\mathfrak{W}$ equals the secrecy capacity of the arbitrarily varying wiretap channel $\overline{\mathfrak{W}}$.*

Proof. The proof was given for an ordinary arbitrarily varying channel AVC without secrecy criterion in [10] and for an AVWC under the weak secrecy criterion in [14]. Let $\tilde{W}_1, \ldots, \tilde{W}_n$ be averaged channels as defined in (11) and a channel $W_{\tilde{q}}^n : A^n \to \mathcal{P}(B^n)$ with $\tilde{q} = \prod_{i=1}^{n} q_i$, $\tilde{q} \in \mathcal{P}(S^n)$, $q_i \in \mathcal{P}(S)$ defined by

$$W_{\tilde{q}}^n(y^n|x^n) = \prod_{i=1}^{n} \tilde{W}_i(y_i|x_i) = \prod_{i=1}^{n} W_{q_i}(y_i|x_i) = \sum_{s^n \in S^n} W^n(y^n|x^n, s^n)\tilde{q}(s^n)$$

If we now use the same (n, J_n) code $\mathcal{C}_n$ defined by the same pair of encoder and decoding sets as for the AVWC $\mathfrak{W}$ the error probability for transmission of a single codeword by the channel $W_{\tilde{q}}^n$ is given by

$$\sum_{x^n \in A^n} E(x^n|j)W_{\tilde{q}}^n(D_j^c|x^n) = \sum_{s^n \in S^n} \tilde{q}(s^n) \sum_{x^n \in A^n} E(x^n|j)W_{s^n}^n(D_j^c|x^n)$$

and we can bound the average error probability by

$$\frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j)W_{\tilde{q}}^n(D_j^c|x^n)$$

$$\leq \max_{s^n \in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j)W_{s^n}^n(D_j^c|x^n) = e(\mathcal{C}_n) \ .$$

Otherwise, because $\mathfrak{W}$ is a subset of $\overline{\mathfrak{W}}^n$, which is the closure of the set of channels $(W_{\tilde{q}}^n, V_{\tilde{q}}^n)$, the opposite inequality holds for the channel $W_{\tilde{q}}^n$ that maximizes the error probability. Because $V_{\tilde{q}}^n$ is defined analogously to $W_{\tilde{q}}^n$, we can define for the (n, J_n) code

$$\hat{V}(z^n|j) := \sum_{x^n \in A^n} E(x^n|j)V_{\tilde{q}}^n(z^n|x^n) \tag{12}$$

for all $z^n \in C^n$, $j \in \mathcal{J}_n$. Then

$$\hat{V}(z^n|j) = \sum_{s^n \in S^n} \tilde{q}(s^n) \sum_{x^n \in A^n} E(x^n|j) V_{s^n}^n(z^n|x^n) = \sum_{s^n \in S^n} \tilde{q}(s^n) \hat{V}_{s^n}^n(z^n|j) \quad (13)$$

and because of the convexity of the mutual information in the channel $\hat{V}$ and (13) it holds that

$$I(J, Z_{\tilde{q}}^n) \leq \sum_{s^n \in S^n} \tilde{q}(s^n) I(J; Z_{s^n}^n) \leq \sup_{s^n} I(J, Z_{s^n}^n). \quad (14)$$

Now because $\{\hat{V}_{s^n}^n(z^n|j) : s^n \in S^n\}$ is a subset of $\{\hat{V}(z^n|j) : \tilde{q} \in \mathcal{P}(S^n)\}$ we end in

$$\sup_{\tilde{q} \in \mathcal{P}(S^n)} I(J, Z_{\tilde{q}}^n) = \sup_{s^n} I(J, Z_{s^n}^n) \ .$$

$\square$

Now we can proceed in the construction of the random code of the AVWC $\mathfrak{W}$.

Definition 4. *We call a channel to the eavesdropper a best channel if there exist a channel $V_{q^*} \in \{V_q : q \in \mathcal{P}(S)\}$ such that all other channels from $\{V_q : q \in \mathcal{P}(S)\}$ are degraded versions of V_{q^*}. If we denote the output of any channel V_q, $q \in \mathcal{P}(S)$ by Z_q it holds that*

$$X \to Z_{q^*} \to Z_q, \quad \forall q \in \mathcal{P}(S). \quad (15)$$

Proposition 1. *Provided that there exist a best channel to the eavesdropper, for the random code secrecy capacity $C_{S,ran}(\mathfrak{W})$ of the AVWC $\mathfrak{W}$ it holds that*

$$C_{S,ran}(\mathfrak{W}) \geq \max_{p \in \mathcal{P}(A)} (\min_{q \in \mathcal{P}(S)} I(p, W_q) - \max_{q \in \mathcal{P}(S)} I(p, V_q)). \quad (16)$$

Proof. The proof is based on Ahlswedes *robustification technique* [3] and is divided in two parts:

step 1): The set

$$\overline{\mathcal{W}} := \{(W_q^n, V_q^n) : q \in \mathcal{P}(S)\}$$

corresponds to a compound wiretap channel indexed by the set of all possible distributions $q \in \mathcal{P}(S)$ on the set of states S. First we show, that there exist a deterministic code for the compound wiretap channel $\overline{\mathcal{W}}$ that achieves the lower bound on the random code secrecy capacity of the AVWC $\mathfrak{W}$ given in (16).

In [8] it was shown that for a compound wiretap channel $\{(W_t, V_t) : t \in \theta\}$ without channel state information at the legitimate receivers the secrecy capacity is bounded from below by

$$C_{S,\text{comp}} \geq \max_{p \in \mathcal{P}(A)} (\min_{t \in \theta} I(p, W_s) - \max_{t \in \theta} I(p, V_s)). \quad (17)$$

In accordance with the proof of (17) in [8] we define a set of i.i.d. random variables $\{X_{jl}\}_{j \in [J_n], l \in [L_n]}$ each according to the distribution $p' \in \mathcal{P}(A^n)$ with

$$p'(x^n) := \begin{cases} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}_{p,\delta}^n)} & \text{if } x^n \in \mathcal{T}_{p,\delta}^n, \\ 0 & \text{otherwise,} \end{cases} \quad (18)$$

for any $p \in \mathcal{P}(A)$, and where J_n and L_n are chosen as

$$J_n = \lfloor 2^{n[\inf_{q \in \mathcal{P}(S)} I(p,W_q) - \sup_{q \in \mathcal{P}(S)} I(p,V_q) - \tau]} \rfloor \qquad (19)$$

$$L_n = \lfloor 2^{n[\sup_{q \in \mathcal{P}(S)} I(p,V_q) + \frac{\tau}{4}]} \rfloor \qquad (20)$$

with $\tau > 0$. Now we assume that there exist a best channel to the eavesdropper V_{q^*} in contrast to the proof in [8]. Hence by the definition of V_{q^*} in (15) and because the mutual Information $I(p,V)$ is convex in V and every member of $\{V_q\}_{q \in \mathcal{P}(S)}$ is a convex combination of the set $\{V_s\}_{s \in S}$, it holds that

$$I(p,V_{q^*}) = \sup_s I(p,V_s) = \sup_{q \in \mathcal{P}(S)} I(p,V_q) \qquad (21)$$

for all $p \in \mathcal{P}(A)$. Note that because of (21) for $|S| < \infty$ $V_{q^*} \in \{V_s : s \in S\}$, which means that q^* is a one-point distribution.

By the definition of the compound channel $\overline{W}$ the channels to the eavesdropper are of the form

$$V_q^n(z^n|x^n) := \prod_{i=1}^n V_q(z_i|x_i) \qquad (22)$$

for all $q \in \mathcal{P}(S)$. Then following the same approach as in the proof in [8] we define

$$\tilde{Q}_{q,x^n}(z^n) = V_q^n(z^n|x^n) \cdot \mathbf{1}_{\mathcal{T}_{V_q,\delta}^n(x^n)}(z^n),$$

and

$$\Theta_q'(z^n) = \sum_{x^n \in \mathcal{T}_{p,\delta}^n} p'(x^n) \tilde{Q}_{q,x^n}(z^n). \qquad (23)$$

for all $z^n \in C^n$. Now let $\mathcal{B} := \{z^n \in C^n : \Theta_q'(z^n) \geq \epsilon\alpha_q\}$ where $\epsilon = 2^{-nc'\delta^2}$ (cf. Lemma 1) and α_q is from (5) in Lemma 2 computed with respect to p and V_q. By Lemma 2 the support of Θ_q' has cardinality $\leq \alpha_q^{-1}$ since for each $x^n \in \mathcal{T}_{p,\delta}^n$ it holds that $\mathcal{T}_{V_q,\delta}^n(x^n) \subset \mathcal{T}_{pV_q,2|A|\delta}^n$, which implies that $\sum_{z^n \in \mathcal{B}} \Theta_q(z^n) \geq 1 - 2\epsilon$, if

$$\Theta_q(z^n) = \Theta_q'(z^n) \cdot \mathbf{1}_{\mathcal{B}}(z^n) \quad \text{and}$$

$$Q_{q,x^n}(z^n) = \tilde{Q}_{q,x^n}(z^n) \cdot \mathbf{1}_{\mathcal{B}}(z^n). \qquad (24)$$

Now it is obvious from (23) and the definition of the set $\mathcal{B}$ that for any $z^n \in \mathcal{B}$ $\Theta_q(z^n) = \mathbb{E}Q_{q,X_{jl}}(z^n) \geq \epsilon\alpha_q$ if $\mathbb{E}$ is the expectation value with respect to the distribution p'. Let β_q defined as in (6) with respect to V_q. For the random variables $\beta_q^{-1}Q_{q,X_{jl}}(z^n)$ define the event

$$\iota_j(q) = \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{q,X_{jl}}(z^n) \in [(1 \pm \epsilon)\Theta_q(z^n)] \right\}, \qquad (25)$$

and keeping in mind that $\Theta_q(z^n) \geq \epsilon\alpha_q$ for all $z^n \in \mathcal{B}$. Then it follows that for all $j \in [J_n]$ and for all $s \in S$

$$\Pr\{(\iota_j(q))^c\} \leq 2|C|^n \exp\left(-L_n \frac{2^{-n[I(p,V_q)+g(\delta)]}}{3}\right) \qquad (26)$$

by Lemma 3, Lemma 2, and our choice $\epsilon = 2^{-nc'\delta^2}$ with $g(\delta) := f_1(\delta) + f_2(\delta) + 3c'\delta^2$. Making $\delta > 0$ sufficiently small we have for all sufficiently large $n \in \mathbb{N}$

$$L_n 2^{-n[I(p,V_q)+g(\delta)]} \geq 2^{n\frac{\tau}{8}}.$$

Thus, for this choice of δ the RHS of (26) is double exponential in n uniformly in $q \in \mathcal{P}(S)$ and can be made smaller than ϵJ_n^{-1} for all $j \in [J_n]$ and all sufficiently large $n \in \mathbb{N}$. I.e.

$$\Pr\{(\iota_j(q))^c\} \leq \epsilon J_n^{-1} \quad \forall q \in \mathcal{P}(S) \tag{27}$$

Now we will show that we can achieve reliable transmission to the legitimate receiver governed by $\{(W_q^n : q \in \mathcal{P}(S)\}$ for all messages $j \in [J_n]$ when randomising over the index $l \in L_n$ but without the need of decoding $l \in [L_n]$. To this end define $\mathcal{X} = \{X_{jl}\}_{j \in [J_n], l \in [L_n]}$ to be the set of random variables with X_{jl} are i.i.d. according to p' defined in (18). Define now the random decoder $\{D_j(\mathcal{X})\}_{j \in [J_n]} \subseteq B^n$ analogously as in [8], [7]. Then it was shown by the authors, that there exist a sequence of (n, J_n) codes for the compound wiretap channel in the particular case without CSI with arbitrarily small mean average error

$$\mathbb{E}_{\mathcal{X}}(\lambda_n^{(q)}(\mathcal{X})) \leq 2^{-na}$$

for all $q \in \mathcal{P}(S)$ and sufficiently large $n \in \mathbb{N}$. Additionally we define for each $q \in \mathcal{P}(S)$

$$\iota_0(q) = \{\lambda_n^{(q)}(\mathcal{X})) \leq 2^{-n\frac{a}{2}}\} \tag{28}$$

and set

$$\iota := \bigcap_{q \in \mathcal{P}_0(n,S)} \bigcap_{j=0}^{J_n} \iota_j(q) \tag{29}$$

Then with (27), (28) and applying the union bound we obtain

$$\Pr\{\iota^c\} \leq 2^{-nc}$$

for a suitable positive constant $c > 0$ and all sufficiently large $n \in \mathbb{N}$ (Cf. [8]).

Hence, we have shown that there exist realisations $\{x_{jl}\}$ of $\{X_{jl}^n\}_{j \in [J_n], l \in [L_n]}$ such that $x_{jl} \in \iota$ for all $j \in [J_n]$ and $l \in [L_n]$. Now following the same argumentation as in [8], [7] we obtain that there is a sequence of (n, J_n) codes that for all codewords $\{x_{jl}\}$ it follows by construction that

$$\frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} W_q^n(D_j^c | x_{jl}) \leq 2^{-na'} \tag{30}$$

is fulfilled for $n \in \mathbb{N}$ sufficiently large and for all $q \in \mathcal{P}(S)$ with $a' > 0$. So we have found a (n, J_n) code with average error probability upper bounded by (30). Further, for the given code and a random variable J uniformly distributed on the message set $\{1, \ldots, J_n\}$ it holds that

$$I(p_J; V_q^n) \leq \epsilon' \tag{31}$$

uniformly in $q \in \mathcal{P}(S)$. Both (30) and (31) ensure that in the scenario of the compound wiretap channel $\overline{\mathcal{W}}$ the legitimate receiver can identify each message j from the message set $\{1, \ldots, J_n\}$ with high probability, while at the same time the eavesdropper receives almost no information about it. That is, that all numbers R_S with

$$R_S \leq \inf_{q \in \mathcal{P}(S)} I(p, W_q) - \sup_{q \in \mathcal{P}(S)} I(p, V_q) \tag{32}$$

are achievable secrecy rates of the compound wiretap channel $\overline{\mathcal{W}}$.

step 2): *Robustification*: In the second step we derive from the deterministic (n, J_n) code for the above mentioned compound wiretap channel $\overline{\mathcal{W}}$ a (n, J_n) random code $\mathcal{C}_n^{\mathrm{ran}}$ for the AVWC $\mathfrak{W}$, which achieves the same secrecy rates. We note first that by (21) and (31)

$$\max_{s^n \in S^n} I(p_J, V_{s^n}) = I(p_J, V_{q^*}^n) \leq \epsilon', \tag{33}$$

which means, that, due to the assumption of a best channel to the eavesdropper, the code achieving the secrecy rate for the best channel to the eavesdropper fulfills the secrecy criterion for a channel with any state sequence $s^n \in S^n$. Now, as already mentioned we use the robustification technique (cf. Lemma 4) to derive from the deterministic code $\mathcal{C}_{\overline{\mathcal{W}}} = \{x_{jl}, D_j : j \in [J_n], l \in [L_n]\}$ of the compound wiretap channel $\overline{\mathcal{W}}$ the random code for the AVWC $\mathfrak{W}$. Therefore, for now let S to be finite. With (30) it holds that

$$\frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} \sum_{s^n \in S^n} W^n(D_j | x_{jl}, s^n) q^{\otimes n}(s^n) \geq 1 - 2^{-na'} \tag{34}$$

for all $q^{\otimes n} = \prod_{i=1}^n q$ and in particular for all $q \in \mathcal{P}_0(n, S)$. Now let $\pi \in \Pi_n$ be the bijection on S^n induced by the permutation $\sigma \in \Sigma_n$. Since (7) is fulfilled with

$$f(s^n) = \frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} W^n(D_j | x_{jl}, s^n) \tag{35}$$

it follows from (8) that

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} \frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} W^n(D_j | x_{jl}, \pi(s^n)) \geq 1 - (n+1)^{|S|} 2^{-na'} \tag{36}$$

for all $s^n \in S^n$. Hence by defining $\mathcal{C}^\pi := \{\pi^{-1}(x_{jl}^n), \pi^{-1}(D_j)\}$ as a member of a family of codes $\{\mathcal{C}^\pi\}_{\pi \in \Pi_n}$ together with a random variable K distributed according to μ as the uniform distribution on Π_n, (36) is equivalent to

$$\mathbb{E}_\mu(\bar{\lambda}_n(\mathcal{C}^K, W_{s^n}^n)) \leq (n+1)^{|S|} 2^{-na'} =: \lambda_n \tag{37}$$

with $\bar{\lambda}_n(\mathcal{C}^\pi, W_{s^n}^n)$ as the respective average error probability for $K = \pi$ and it holds for all $s^n \in S^n$. Thus we have shown that

$$\mathcal{C}_n^{\mathrm{ran}} := \{(\pi^{-1}(x_{jl}), \pi^{-1}(D_j)) : j \in [J_n], l \in [L_n], \pi \in \Pi_n, \mu\} \tag{38}$$

is a (n, J_n, Π_n, μ) random code for the AVC channel $\mathcal{W}^n = \{W_{s^n} : s^n \in S^n\}$ with the mean average error probability $\mathbb{E}_\mu(\bar{\lambda}_n(\mathcal{C}^K, W_{s^n}^n))$ upper bounded by λ_n as in (37).

Now it is easily seen that

$$p_{JZ_{q*}^n}^{\mathcal{C}^\pi}(j, z^n) = \frac{1}{J_n} \frac{1}{L_n} \sum_{l=1}^{L_n} V_{q*}^n(\pi^{-1}(z^n)|\pi^{-1}(x_{jl})) = p_{JZ_{q*}^n} . \tag{39}$$

Actually, it still holds that

$$p_{JZ_{q*}^n}^{\mathcal{C}^r}(j, z^n) = \frac{1}{n!} \sum_{\pi \in \Pi_n} p_{JZ_{q*}^n}^{\mathcal{C}^\pi}(j, z^n) = p_{JZ_{q*}^n} . \tag{40}$$

With (39) and the representation of the mutual information by the information divergence we obtain from (33)

$$\mathbb{E}_\mu(D(p_{JZ_{q*}^n}^{\mathcal{C}^K} || p_J \otimes p_{Z_{q*}^n}^{\mathcal{C}^K})) = \frac{1}{n!} \sum_{\pi \in \Pi_n} D(p_{JZ_{q*}^n}^{\mathcal{C}^\pi} || p_J \otimes p_{Z_{q*}^n}^{\mathcal{C}^\pi})$$

$$= \frac{1}{n!} \sum_{\pi \in \Pi_n} D(p_{JZ_{q*}^n} || p_J \otimes p_{Z_{q*}^n}) = I(p_J, V_{q*}^n) \leq \epsilon' . \tag{41}$$

Thus we have constructed a random (n, J_n, Γ, μ) code $\mathcal{C}_n^{\mathrm{ran}}$ with mean average error probability bounded for all $s^n \in S^n$ as in (37) and which fulfills the strong secrecy criterion almost surely, provided that there exist a best channel to the eavesdropper. By the construction of the random code it follows that the secrecy rates given by (32) for the compound wiretap channel $\overline{\mathcal{W}}$ achieved by the deterministic code $\mathcal{C}_{\overline{\mathcal{W}}}$ are achievable secrecy rates for the AVWC $\mathfrak{W}$ with random code $\mathcal{C}_n^{\mathrm{ran}}$. That is, we have shown that all rates R_S with

$$R_S \leq \max_{p \in \mathcal{P}(A)} (\min_{q \in \mathcal{P}(S)} I(p, W_q) - \max_{q \in \mathcal{P}(S)} I(p, V_q)) . \tag{42}$$

are achievable secrecy rates of the arbitrarily varying wiretap channel AVWC with random code $\mathcal{C}_n^{\mathrm{ran}}$. $\square$

3.3   Deterministic Code Construction

Because the code $\mathcal{C}^\pi$ that is used for the transmission of a single message is subjected to a random selection, reliable transmission can only be guaranteed if the outcome of the random experiment can be shared by both the transmitter and the receiver. One way to inform the receiver about the code that is chosen is to add a short prefix to the actual codeword. Provided that the number of codes is small enough, the transmission of these additional prefixes causes no essential loss in rate. In the following we use the *elimination technique* by Ahlswede [2] which has introduced the above approach to derive deterministic codes from random codes for determining capacity of arbitrarily varying channels. Temporarily we drop the requirement of a best channel to the eavesdropper and state the following theorem.

Theorem 2. *1. Assume that for the AVWC $\mathfrak{W}$ it holds that $C_{S,\mathrm{ran}}(\mathfrak{W}) > 0$. Then the secrecy capacity $C_S(\mathfrak{W})$ equals its random code secrecy capacity $C_{S,\mathrm{ran}}(\mathfrak{W})$,*

$$C_S(\mathfrak{W}) = C_{S,\mathrm{ran}}(\mathfrak{W}), \tag{43}$$

if and only if the channel to the legitimate receiver is non-symmetrisable.
2. If $C_{S,\mathrm{ran}}(\mathfrak{W}) = 0$ it always holds that $C_S(\mathfrak{W}) = 0$.

First, if the channel to the legitimate receiver is symmetrisable then the deterministic code capacity of the channel to the legitimate receiver equals zero by Theorem 1 and no reliable transmission of messages is possible. Hence the deterministic code secrecy capacity of the arbitrarily varying wiretap channel also equals zero although the random code secrecy capacity could be greater than zero. So we can restrict to the case in which the channel to the legitimate receiver is non-symmetrisable. If $C_S(\mathfrak{W}) = C_{S,\mathrm{ran}}(\mathfrak{W}) > 0$, then the channel to the legitimate receiver must be nonsymmetrisable. For the other direction, because the secrecy capacity of the AVWC $\mathfrak{W}$ cannot be greater than the random code secrecy capacity it suffices to show that $C(\{W_{s^n}\}) > 0$ implies that $C_S(\mathfrak{W}) \geq C_{S,\mathrm{ran}}(\mathfrak{W})$. Here $C(\{W_{s^n}\})$ denotes the capacity of the arbitrarily varying channels to the legitimate receiver without secrecy. The proof is given in the two paragraphs *Random code reduction* and *Elimination of randomness*.

Random Code Reduction. We first reduce the random code $\mathcal{C}^{\mathrm{ran}}$ to a new random code selecting only a small number of deterministic codes from the former, and averaging over this codes gives a new random code with a constant small mean average error probability, which additionally fulfills the secrecy criterion.

Lemma 6. *(Random Code Reduction) Let $\mathcal{C}(\mathcal{Z})$ be a random code for the AVWC $\overline{\mathfrak{W}}$ consisting of a family $\{\mathcal{C}(\gamma)\}_{\gamma \in \Gamma}$ of wiretap codes where γ is chosen according to the distribution μ of $\mathcal{Z}$. Then let*

$$\bar{e}(\mathcal{C}_n^{\mathrm{ran}}) = \max_{s^n} \mathbb{E}_\mu e(s^n | \mathcal{C}(\mathcal{Z})) \leq \lambda_n \quad and, \quad \max_{s^n} \mathbb{E}_\mu I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z})) \leq \epsilon_n' \ . \tag{44}$$

Then for any ϵ and K satisfying

$$\epsilon > 4 \max\{\lambda_n, \epsilon_n'\} \quad and \quad K > \frac{2n \log |A|}{\epsilon}(1 + n \log |S|) \tag{45}$$

there exist K deterministic codes $\mathcal{C}_i$, $i = 1, \ldots, K$ chosen from the random code by random selection such that

$$\frac{1}{K} \sum_{i=1}^{K} e(s^n | \mathcal{C}_i) \leq \epsilon \quad and \quad \frac{1}{K} \sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}_i) \leq \epsilon \tag{46}$$

for all $s^n \in S^n$.

Proof. The proof is analogue to the proof of Lemma 6.8 [10], where a similar assertion in terms of the maximal probability of error for single user AVCs without secrecy criterion is established. Cf. also [2]. Let $\mathcal{Z}$ be the random variable distributed according to μ on Γ for the (n, J_n, Γ, μ) random code. Now consider K independent repetitions of the random experiment of code selections according to μ and call the according random variables $\mathcal{Z}_i$, $i \in \{1, \ldots, K\}$. Then for any $s^n \in S^n$ it holds that

$$
\Pr\Big\{ \frac{1}{K} \sum_{i=1}^K e(s^n | \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \quad \text{or} \quad \frac{1}{K} \sum_{i=1}^K I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \Big\}
$$

$$
\leq \Pr\Big\{ \exp \sum_{i=1}^K \frac{e(s^n | \mathcal{C}(\mathcal{Z}_i))}{n \log |A|} \geq \exp \frac{K\epsilon}{n \log |A|} \Big\}
$$

$$
+ \Pr\Big\{ \exp \sum_{i=1}^K \frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i))}{n \log |A|} \geq \exp \frac{K\epsilon}{n \log |A|} \Big\},
$$

and by Markov's inequality

$$
\Pr\Big\{ \frac{1}{K} \sum_{i=1}^K e(s^n | \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \quad \text{or} \quad \frac{1}{K} \sum_{i=1}^K I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \Big\}
$$

$$
\leq \exp\Big(-\frac{K\epsilon}{n \log |A|} \Big) \mathbb{E} \exp \sum_{i=1}^K \frac{e(s^n | \mathcal{C}(\mathcal{Z}_i))}{n \log |A|}
$$

$$
+ \exp\Big(-\frac{K\epsilon}{n \log |A|} \Big) \mathbb{E} \exp \sum_{i=1}^K \frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i))}{n \log |A|} \ .
$$

Now because of the independency of the random variables $\mathcal{Z}_i$ and because all $\mathcal{Z}_i$ are distributed as $\mathcal{Z}$ and we have $\exp t \leq 1 + t$, for $0 \leq t \leq 1$ (exp to the base 2), we can give the following upper bounds

$$
\Big(\mathbb{E} \exp \frac{e(s^n | \mathcal{C}(\mathcal{Z}))}{n \log |A|} \Big)^K \leq \Big(1 + \mathbb{E} \frac{e(s^n | \mathcal{C}(\mathcal{Z}))}{n \log |A|} \Big)^K \leq \Big(1 + \frac{\lambda_n}{n \log |A|} \Big)^K \tag{47}
$$

and

$$
\Big(\mathbb{E} \exp \frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}))}{n \log |A|} \Big)^K \leq \Big(1 + \mathbb{E} \frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}))}{n \log |A|} \Big)^K \leq \Big(1 + \frac{\epsilon_n'}{n \log |A|} \Big)^K \ .
$$

$$
\tag{48}
$$

Hence we obtain for any $s^n \in S^n$

$$\Pr\Big\{\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \quad \text{or} \quad \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon\}$$

$$\leq \exp\Big[-K\Big(\frac{\epsilon}{n\log|A|} - \log(1 + \frac{\lambda_n}{n\log|A|})\Big)\Big]$$

$$+ \exp\Big[\Big(-K\big(\frac{\epsilon}{n\log|A|} - \log(1 + \frac{\epsilon'_n}{n\log|A|})\big)\Big)\Big]$$

$$\leq 2\exp\Big[-K\Big(\frac{\epsilon}{n\log|A|} - \log(1 + \max\{\frac{\lambda_n}{n\log|A|}, \frac{\epsilon'_n}{n\log|A|}\})\big)\Big] \ .$$

Then

$$\Pr\Big\{\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}(\mathcal{Z}_i)) \leq \epsilon \text{ and } \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \leq \epsilon, \forall s^n \in S^n\Big\}$$

$$\geq 1 - 2|S|^n \exp\Big[-K\Big(\frac{\epsilon}{n\log|A|} - \log(1 + \max\{\frac{\lambda_n}{n\log|A|}, \frac{\epsilon'_n}{n\log|A|}\})\big)\Big], \tag{49}$$

which is strictly positive, if we choose

$$\epsilon \geq 2n\log|A|\log(1 + \max\{\frac{\lambda_n}{n\log|A|}, \frac{\epsilon'_n}{n\log|A|}\})$$

and

$$K \geq \frac{2\log|A|}{\epsilon}(n + n^2 \log|S|) \ . \tag{50}$$

Now because for $0 \leq t \leq 1$ and log to the base 2 it holds that

$$t \leq \log(1 + t) \leq 2t \ ,$$

we increase the lower bound for choosing ϵ if

$$\epsilon \geq 4\max\{\lambda_n, \epsilon'_n\} \ .$$

and with (50) the assertion of (49) still holds. Hence, we have shown that there exist K realisations $\mathcal{C}_i := \mathcal{C}(\mathcal{Z}_i = \gamma_i)$, $\gamma_i \in \Gamma$, $i \in \{1, \ldots, K\}$ of the random code, which build a new reduced random code with uniform distribution on these codes with mean average error probability and mean secrecy criterion fulfilled by (46). $\qquad \square$

Now, if we assume that the channel to the legitimate receiver is non-symmetrisable, which means that $C(\{W_{s^n}\}) > 0$, and that there exist a random code $\mathcal{C}_n^{\text{ran}}$ that achieves the random code capacity $C_{S,\text{ran}}(\mathfrak{W}) > 0$, then there exist a sequence of random (n, J_n) codes with

$$\lim_{n\to\infty} \max_{s^n \in S^n} \frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{\gamma\in\Gamma}\sum_{x^n\in A^n} E^\gamma(x^n|j) \cdot W_{s^n}^n((D_j^\gamma)^c|x^n)\mu(\gamma) = 0 \ ,$$

$$\liminf_{n\to\infty} \frac{1}{n}\log J_n \to C_{S,\mathrm{ran}}(\mathfrak{W}) > 0,$$

and

$$\lim_{n\to\infty} \max_{s^n\in S^n} \sum_{\gamma\in\Gamma} I(p_J; V_{s^n}^n; \mathcal{C}(\gamma))\mu(\gamma) = 0. \tag{51}$$

Then on account of the random code reduction lemma there exist a sequence of random (n, J_n) codes consisting only of n^3 deterministic codes (cf. (45)) chosen from the former random code, and it holds for any $\epsilon > 0$ and sufficiently large n that

$$\max_{s^n\in S^n} \frac{1}{J_n}\sum_{j=1}^{J_n} \frac{1}{n^3}\sum_{i=1}^{n^3} \sum_{x^n\in A^n} E^i(x^n|j)W_{s^n}^n((D_j^i)^c|x^n) \le \epsilon \tag{52}$$

and

$$\max_{s^n\in S^n} \frac{1}{n^3}\sum_{i=1}^{n^3} I(p_J; V_{s^n}^n; \mathcal{C}_i) \le \epsilon, \tag{53}$$

where $\mathcal{C}_i = \{(E_j^i, D_j^i), j \in \mathcal{J}_n\}$, $i = 1,\ldots,n^3$, and E^i is the stochastic encoder of the deterministic wiretap code. Then the reduced random code consists of the family of codes $\{\mathcal{C}_i\}_{i\in\{1,\ldots,n^3\}}$ together with the uniform distribution $\mu'(i) = \frac{1}{n^3}$ for all $i \in \{1,\ldots,n^3\}$.

Elimination of Randomness. (Cf. Theorem 6.11 in [10])
Now if there exist a deterministic code and $C(\{W_{s^n}\}) > 0$ then there exist a code

$$\{x_i^{k_n}, F_i \subset B^{k_n} : i = 1,\ldots n^3\} \tag{54}$$

where $x_i^{k_n}$ is chosen according to an encoding function $f_i : \{1,\ldots,n^3\} \to A^{k_n}$ with $\frac{k_n}{n} \to 0$ as $n \to \infty$ with error probability

$$\frac{1}{n^3}\sum_{i=1}^{n^3} W^{k_n}(F_i^c|x_i^{k_n}, s^{k_n}) \le \epsilon \tag{55}$$

for any $\epsilon > 0$ and sufficiently large n (cf. (52)) for all $s^{k_n} \in S^{k_n}$. If we now compose a new deterministic code for the AVWC $\mathfrak{W}$ by prefixing the codewords of each C_i

$$\{f_i E_j^i, F_i \times D_j^i : i = 1,\ldots,n^3, j \in [J_n]\} =: \mathcal{C} , \tag{56}$$

the decoder is informed of which encoder E^i is in use for the actual message j if he identifies the prefix correctly. Note that for the transmission of the prefix only the reliability is of interest, because it contains no information about the message $j \in \mathcal{J}_n$ to be sent. Now the new codewords has a length of $k_n + n$, transmit a message from $\{1,\ldots,n^3\} \times \mathcal{J}_n$, where the channel which is determined by the state sequence $s^{k_n+n} \in S^{k_n+n}$ yields an average error probability of

$$\bar{\lambda}_n(\mathcal{C}, W_{s^{k_n+n}}^{(k_n+n)}) \leq \frac{1}{n^3 J_n} \sum_{i=1}^{n^3} \sum_{j\in[J_n]} (\lambda_i + \lambda_j(i))$$

$$\leq \frac{1}{n^3} \sum_{i=1}^{n^3} \lambda_i + \frac{1}{n^3} \sum_{i=1}^{n^3} e_n(s^n, \mathcal{C}_i) \leq 2\epsilon. \tag{57}$$

Here, for each $s^{k_n} \in S^{k_n}$ λ_i means the error probability for transmitting i from $\{1, \ldots, n^3\}$ encoded in $x_i^{k_n}$ by $W_{s^{k_n}}^{k_n}$ followed by the transmission of j, where the codeword is chosen according to the stochastic encoder E_j^i, over the last n channel realisations determined by s^n with error probability $\lambda_j(i)$. This construction is possible due to the memorylessness of the channel.

Now if we turn to the security part of the transmission problem it is easily seen that

$$p_{JZ_{s^{k_n+n}}^{k_n+n}}^{\mathcal{C}}(j, z^{k_n+n}) = \frac{1}{J_n} \frac{1}{n^3} \sum_{i=1}^{n^3} V_{s^{k_n}}^{k_n}(\hat{z}^{k_n}|x_i^{k_n}) \sum_{x^n} E^i(x^n|j) V_{s^n}^n(z^n|x^n)$$

$$= \frac{1}{n^3} \sum_{i=1}^{n^3} V_{s^{k_n}}^{k_n}(\hat{z}^{k_n}|x_i^{k_n}) \cdot p_{JZ_{s^n}^n}^{\mathcal{C}_i}, \tag{58}$$

where $\hat{z}^{k_n}$ are the first k_n components of z^{k_n+n}. With (58) and the representation of the mutual information by the information divergence we obtain that

$$D(p_{JZ_{s^{k_n+n}}^{k_n+n}}^{\mathcal{C}} \| p_J \otimes p_{Z_{s^{k_n+n}}^{k_n+n}}^{\mathcal{C}})$$

$$= D\left(\frac{1}{n^3} \sum_{i=1}^{n^3} V_{s^{k_n}}^{k_n}(\hat{z}^{k_n}|x_i^{k_n}) p_{JZ_{s^n}^n}^{\mathcal{C}_i} \,\Big\|\, \frac{1}{n^3} \sum_{i=1}^{n^3} V_{s^{k_n}}^{k_n}(\hat{z}^{k_n}|x_i^{k_n}) p_J \otimes p_{Z_{s^n}^n}^{\mathcal{C}_i}\right)$$

$$\leq \frac{1}{n^3} \sum_{i=1}^{n^3} D\left(V_{s^{k_n}}^{k_n}(\hat{z}^{k_n}|x_i^{k_n}) p_{JZ_{s^n}^n}^{\mathcal{C}_i} \,\Big\|\, V_{s^{k_n}}^{k_n}(\hat{z}^{k_n}|x_i^{k_n}) p_J \otimes p_{Z_{s^n}^n}^{\mathcal{C}_i}\right)$$

$$= \frac{1}{n^3} \sum_{i=1}^{n^3} D\left(p_{JZ_{s^n}^n}^{\mathcal{C}_i} \,\big\|\, p_J \otimes p_{Z_{s^n}^n}^{\mathcal{C}_i}\right) = \frac{1}{n^3} \sum_{i=1}^{n^3} I(p_J, V_{s^n}^n; \mathcal{C}_i) \leq \epsilon \tag{59}$$

for all $s^n \in S^n$ and $n \in \mathbb{N}$ sufficiently large, where the first inequality follows because for two probability distributions p, q the relative entropy $D(p\|q)$ is a convex function in the pair (p, q) and the last inequality follows by the random code reduction lemma.

Because $\frac{k_n}{n} \to 0$ as $n \to \infty$

$$\lim_{n\to\infty} \frac{1}{k_n + n} \log(n^3 J_n) = \lim_{n\to\infty} (\frac{1}{n} \log J_n + \frac{1}{n} \log(n^3)) = \lim_{n\to\infty} \frac{1}{n} \log J_n , \quad (60)$$

$\mathcal{C}_n$ is a deterministic (n, J_n) code which achieves the same rates as the random code $\mathcal{C}_n^{\text{ran}}$ and so the random code capacity $C_{S,\text{ran}}$ as given in (51), provided that the channel to the legitimate receiver is non-symmetrisable.

Thus, with $\{1, \ldots, J_n\}$ as the message set, $\mathcal{C}_n$ is a deterministic $(n + o(n), n^3 \cdot J_n)$ code with average error probability bounded for all $s^{k_n+n} \in S^{k_n+n}$ as in (57) and which fulfills the strong secrecy criterion as in (59), and which achieves the random code secrecy capacity $C_{S,\mathrm{ran}}$ of the arbitrarily varying wiretap channels AVWC $\mathcal{W}$ which implies that $C_S = C_{S,\mathrm{ran}}$. This concludes the proof.

Note that in the case in which the channel to the legitimate receiver is non-symmetrisable and we know that the deterministic code secrecy capacity $C_S(\mathfrak{W})$ equals zero we can conclude that the random code secrecy capacity $C_{S,\mathrm{ran}}(\mathfrak{W})$ equals zero. As a consequence of the theorem we can state the following assertion.

Corollary 1. *The deterministic code secrecy capacity of the arbitrarily varying wiretap channel* $\mathfrak{W}$*, provided that there exists a best channel to the eavesdropper and under the assumption that the channel to the legitimate receiver is non-symmetrisable, is lower bounded by*

$$C_S(\mathfrak{W}) \geq \max_{p \in \mathcal{P}(A)} \left(\min_{q \in \mathcal{P}(S)} I(p, W_q) - \max_{q \in \mathcal{P}(S)} I(p, V_q) \right) .$$

Proof. Combine the assertions of Proposition 1 and Theorem 2. $\qquad\square$

3.4 Upper Bound on the Capacity of the AVWC $\mathfrak{W}$ and a Multi-letter Coding Theorem

In this section we give an upper bound on the secrecy capacity of the AVWC $\mathfrak{W}$ which corresponds to the bound for the compound wiretap channel built by the same family of channels. In addition we give the proof of the multi-letter converse of the AVWC $\mathfrak{W}$.

Theorem 3. *The secrecy capacity of the arbitrarily varying wiretap channel AVWC* $\mathfrak{W}$ *is upper bounded,*

$$C_S(\mathfrak{W}) \leq \min_{q \in \mathcal{P}(S)} \max_{U \to X \to (YZ)_q} (I(U, Y_q) - I(U, Z_q)) . \tag{61}$$

Proof. By Lemma 5 the capacity of the AVWC $\mathfrak{W}$ equals the capacity of the AVWC $\overline{\mathfrak{W}}$. Obviously, the set $\overline{\mathcal{W}} = \{(W_q^{\otimes n}, V_q^{\otimes n}) : q \in \mathcal{P}(S)\}$ which describes a compound wiretap channel is a subset of $\overline{\mathfrak{W}}^n = \{(W_{\tilde{q}}^n, V_{\tilde{q}}^n) : \tilde{q} \in \mathcal{P}(S^n), \tilde{q} = \prod_{i=1}^n q_i\}$. Now, because we can upper bound the secrecy capacity of the AVWC $\overline{\mathfrak{W}}$ by the secrecy capacity of the worst wiretap channel in the family $\overline{\mathfrak{W}}^n$, together with the foregoing we can upper bound it by the capacity of the worst channel of the compound channel $\overline{\mathcal{W}}$. Hence,

$$C_S(\mathfrak{W}) = C_S(\overline{\mathfrak{W}}) \leq \inf_{\tilde{q}} C_S((W_{\tilde{q}}^n, V_{\tilde{q}}^n))$$

$$\leq \inf_{q} C_S((W_q^n, V_q^n)) = \inf_{q} C_S(W_q, V_q) ,$$

The minimum is attained because of the continuity of $C_S(W_q, V_q)$ on the compact set $\overline{\mathfrak{W}}$. $\qquad\square$

Remark 1. *Consider the special case of an AVWC $\mathfrak{W} = \{(W_{s^n}, V_{r^n}) : s^n \in S_1^n,\ r^n \in S_2^n\}$, where both the state of the main channel $s \in S_1$ and the state of the eavesdropper's channel $r \in S_2$ in every time step can be chosen independently. In addition let us assume that there exist a channel $W_{q_1^*} \in \{W_{q_1} : q_1 \in \mathcal{P}(S_1)\}$, which is a degraded version of all other channels from $\{W_{q_1} : q_1 \in \mathcal{P}(S_1)\}$, and a best channel to the eavesdropper $V_{q_2^*}$ from the set $\{V_{q_2} : q_2 \in \mathcal{P}(S_2)\}$ (cf. Definition 4). Then in accordance with Section 3.5 of [8] the lower bound on the secrecy capacity given in Corollary 1 matches the upper bound from Theorem 3. Thus we can conclude that under the assumption, that the channel to legitimate receiver is non-symmetrisable, the capacity of the AVWC $\mathfrak{W}$ is given by*

$$C_S(\mathfrak{W}) = \max_{U \to X \to (Y_{q_1^*} Z_{q_2^*})} \left(I(U, Y_{q_1^*}) - I(U, Y_{q_2^*}) \right) .$$

Now in addition to Theorem 3 we give a multi-letter formula of the upper bound of the secrecy rates. Therefore we need the following lemma used in analogy to Lemma 3.7 in [8].

Lemma 7. *For the arbitrarily varying wiretap channel AVWC $\mathfrak{W}^n$ the limit*

$$\lim_{n \to \infty} \frac{1}{n} \max_{U \to X^n \to (Y^n Z^n)_{\tilde{q}}} \left(\inf_{\tilde{q} \in \mathcal{P}(S^n)} I(U, Y_{\tilde{q}}^n) - \sup_{\tilde{q} \in \mathcal{P}(S^n)} I(U, Z_{\tilde{q}}^n) \right)$$

exists.

The proof is carried out in analogy to Lemma 3.7 in [8] and therefore omitted.

Theorem 4. *The secrecy capacity of the arbitrarily varying wiretap channel AVWC $\mathfrak{W}$ is upper bounded by*

$$C_S(\mathfrak{W}) \leq \lim_{n \to \infty} \frac{1}{n} \max_{U \to X^n \to (Y^n Z^n)_{\tilde{q}}} \left(\inf_{\tilde{q} \in \mathcal{P}(S^n)} I(U, Y_{\tilde{q}}^n) - \sup_{\tilde{q} \in \mathcal{P}(S^n)} I(U, Z_{\tilde{q}}^n) \right) , \quad (62)$$

where $\tilde{q} = \prod_{i=1}^n q_i$, $q_i \in \mathcal{P}(S)$ and $Y_{\tilde{q}}^n, Z_{\tilde{q}}^n$ are the outputs of the channels $W_{\tilde{q}}^n$ and $V_{\tilde{q}}^n$ respectively.

Proof. Let $(\mathcal{C}_n)_{n \in \mathbb{N}}$ be any sequence of (n, J_n) codes such that with

$$\sup_{s^n \in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j) W_{s_n}^n (D_j^c|x^n) =: \varepsilon_{1,n} \text{ and, } \sup_{s^n \in S^n} I(J, Z_{s^n}^n) =: \varepsilon_{2,n}$$

it holds that $\lim_{n \to \infty} \varepsilon_{1,n} 0 =$ and $\lim_{n \to \infty} \varepsilon_{2,n}$, where J denotes the random variable which is uniformly distributed on the message set $\mathcal{J}_n$. Because of Lemma 5 we obtain that for the same sequences of (n, J_n) codes

$$\lim_{n \to \infty} \sup_{\tilde{q} \in \mathcal{P}(S^n)} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j) W_{\tilde{q}}^n (D_j^c|x^n) = \lim_{n \to \infty} \varepsilon_{1,n} = 0 \quad (63)$$

and

$$\lim_{n \to \infty} \sup_{\tilde{q} \in \mathcal{P}(S^n)} I(J, Z_{\tilde{q}}^n) = \lim_{n \to \infty} \varepsilon_{2,n} = 0 \ . \tag{64}$$

Now let us denote another random variable by $\hat{J}$ with values in $\mathcal{J}_n$ determined by the Markov chain $J \to X^n \to Y_{\tilde{q}}^n \to \hat{J}$, where the first transition is governed by E, the second by $W_{\tilde{q}}^n$, and the last by the decoding rule. Now the proof is analogue to the proof of Proposition 3.8 in [8]. For any $\tilde{q} \in \mathcal{P}(S^n)$ we have from data processing and Fano's inequality

$$(1 - \varepsilon_{1,n}) \log J_n \leq I(J, Y_{\tilde{q}}^n) + 1.$$

We then use the validity of the secrecy criterion (64) to derive

$$(1 - \varepsilon_{1,n}) \log J_n \leq I(J, Y_{\tilde{q}}^n) - \sup_{\tilde{q}} I(J, Z_{\tilde{q}}^n) + \varepsilon_{2,n} + 1$$

for any $\tilde{q} \in \mathcal{P}(S^n)$. Since the LHS does not depend on $\tilde{q}$ we end in

$$(1 - \varepsilon_{1,n}) \log J_n \leq \max_{U \to X^n \to Y_{\tilde{q}}^n Z_{\tilde{q}}^n} (\inf_{\tilde{q}} I(U, Y_{\tilde{q}}^n) - \sup_{\tilde{q}} I(U, Z_{\tilde{q}}^n)) + \varepsilon_{2,n} + 1 \ .$$

Dividing by $n \in \mathbb{N}$ and taking $\overline{\lim}$ concludes the proof.   □

Now if we consider the set $\overline{\mathcal{W}} = \{(W_q^{\otimes n}, V_q^{\otimes n}) : q \in \mathcal{P}(S)\}$ as a subset of $\overline{\mathfrak{W}}^n = \{(W_{\tilde{q}}^n, V_{\tilde{q}}^n) : \tilde{q} \in \mathcal{P}(S^n), \tilde{q} = \prod_{i=1}^n q_i\}$ and the same sequence $(\mathcal{C}_n)_{n \in \mathbb{N}}$ of (n, J_n) codes for the AVWC $\mathfrak{W}$ for which (63) and (64) holds, we can conclude that

$$\lim_{n \to \infty} \sup_{q \in \mathcal{P}(S)} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j) W_q^{\otimes n}(D_j^c|x^n) \leq \lim_{n \to \infty} \varepsilon_{1,n} \tag{65}$$

and

$$\lim_{n \to \infty} \sup_{q \in \mathcal{P}(S)} I(J, Z_q^n) \leq \lim_{n \to \infty} \varepsilon_{2,n} \ , \tag{66}$$

with $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ as above. Then we can conclude with the same argumentation as in the previous proof,

Corollary 2. *The secrecy capacity of the arbitrarily varying wiretap channel AVWC $\mathfrak{W}$ is upper bounded by*

$$C_S(\mathfrak{W}) \leq \lim_{n \to \infty} \frac{1}{n} \max_{U \to X^n \to (Y^n Z^n)_q} (\inf_{q \in \mathcal{P}(S)} I(U, Y_q^n) - \sup_{q \in \mathcal{P}(S)} I(U, Z_q^n)) \ ,$$

where $q \in \mathcal{P}(S)$ and Y_q^n, Z_q^n are the outputs of the channels $W_q^{\otimes n}$ and $V_q^{\otimes n}$ respective.

Now, using standard arguments concerning the use of the channels defined by $P_{Y_q|U} = W_q \cdot P_{X|U}$ and $P_{Z_q|U} = V_q \cdot P_{X|U}$ instead of W_q and V_q and applying the assertion of Corollary 1 to the n-fold product of channels W_q and V_q, we are able to give the coding theorem for the multi-letter case of the AVWC with a best channel to the eavesdropper.

Theorem 5. *Provided that there exist a best channel to the eavesdropper, the multi-letter expression for the secrecy capacity $C_S(\mathfrak{W})$ of the AVWC $\mathfrak{W}$ is given by*

$$C_S(\mathfrak{W}) = \lim_{n \to \infty} \frac{1}{n} \max_{U \to X^n \to (Y^n Z^n)_q} \left(\inf_{q \in \mathcal{P}(S)} I(U, Y_q^n) - \sup_{q \in \mathcal{P}(S)} I(U, Z_q^n) \right) ,$$

if the channel to the legitimate receiver is non-symmetrisable, and is zero otherwise.

References

1. Ahlswede, R.: A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to shannon's zero error capacity. The Annals of Mathematical Statistics 41(3), 1027–1033 (1970)
2. Ahlswede, R.: Elimination of correlation in random codes for arbitrarily varying channels. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 44, 159–175 (1978)
3. Ahlswede, R.: Arbitrarily varying channels with states sequence known to the sender. IEEE Trans. on Inf. Theory 32(5), 621–629 (1986)
4. Ahlswede, R., Csiszár, I.: Common randomness in information theory and cryptography-part I: Secret sharing. IEEE Trans. on Inf. Theory 39(4), 1121–1132 (1993)
5. Ahlswede, R., Csiszár, I.: Common randomness in information theory and cryptography-part ii: Cr capacity. IEEE Trans. on Inf. Theory 44(1), 225–240 (1998)
6. Ahlswede, R., Wolfowitz, J.: The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet. Z. Wahrscheinlichkeitstheorie verw. Gebiete 15, 186–194 (1970)
7. Bjelacović, I., Boche, H., Sommerfeld, J.: Capacity results for compound wiretap channels. In: Proc. IEEE Information Theory Workshop, pp. 60–64 (2011)
8. Bjelacović, I., Boche, H., Sommerfeld, J.: Secrecy results for compound wiretap channels. Submitted to Problems of Information Transmission (2011), http://arxiv.org/abs/1106.2013v1
9. Bloch, M., Laneman, J.: On the secrecy capacity of arbitrary wiretap channel. In: Forty-Sixth Annual Allerton Conference, Allerton House, Illinois, USA (2008)
10. Csiszár, I., Körner, J.: Information Theory: Coding Theorems for Discrete Memoryless Systems. Akademiai Kiado (1981)
11. Csiszár, I., Narayan, P.: The capacity of the arbitrarily varying channel revisited: positivity, constraints. IEEE Trans. on Inf. Theory 34(2), 181–193 (1988)
12. Ericson, T.: Exponential error bounds for random codes in the arbitrarily varying channel. IEEE Trans. on Inf. Theory 31(1), 42–48 (1985)

13. Liang, Y., Kramer, G., Poor, H., Shamai, S.: Compound Wiretap Channels. EURASIP Journal on Wireless Communications and Networking (2008)
14. MolavianJazi, E.: Secure Communications over Arbitrarily Varying Wiretap Channels. Master's thesis, Graduate School of the University of Notre Dame (2009)
15. Shannon, C.: The zero error capacity of a noisy channel. IRE Trans. Information Theory IT-2, 8–19 (1956)
16. Wyrembelski, R.F., Bjelaković, I., Oechtering, T.J., Boche, H.: Optimal coding strategies for bidirectional broadcast channels under channel uncertainty. IEEE Trans. on Communications 58(10), 2984–2994 (2010)