OPTIMAL TRANSCEIVER DESIGN FOR WIRETAP CHANNELS WITH SIDE INFORMATION

Holger Boche and Rafael F. Schaefer

Lehrstuhl für Theoretische Informationstechnik
Technische Universität München, Germany

ABSTRACT

The wiretap channel models the communication scenario where two legitimate users want to communicate in such a way that an external wiretapper is kept ignorant. In this paper, the *wiretap channel with side information* is studied, where the wiretapper has additional side information about the transmitted message available for post-processing. The secrecy of the message is modeled by the decoding performance of the wiretapper. It is required for the wiretapper to have worst behavior of decoding performance regardless of the decoding strategy that is used. The secrecy capacity is derived and shown to be equal to the one of the classical wiretap channel without side information available at the wiretapper. Further, the corresponding optimal transceiver design is characterized.

Index Terms— Wiretap channel, strong secrecy, side information, secrecy capacity, optimal transceiver design.

1. INTRODUCTION

The concept of physical layer security is becoming more and more attractive, since it solely uses the physical properties of a wireless channel to establish security. So, regardless of the applied post-processing at non-legitimate receivers, the confidential information cannot be reproduced from the received signal with arbitrarily high probability. Recently, there is growing interest in physical layer security; for instance see [1–4] and references therein.

Physical layer security was initiated by Wyner, who introduced the *wiretap channel* [5]. It describes the simplest scenario involving security with one legitimate transmitter-receiver pair and one external wiretapper to be kept secret. The aim is to encode and transmit the message in such a way that the legitimate receiver is able to decode the message and, at the same time, the wiretapper is prevented to infer the confidential information from the received signal. The wiretap channel is widely studied under several aspects, cf. [5–15].

All these works have in common that the wiretapper has only the received channel output available for post-processing. Here we study the *wiretap channel with side information*, where we consider a wiretapper which has additional side information about the transmitted message available. This models a priori knowledge about the transmitted message which allows the wiretapper to restrict the message to a certain subset of all possible messages. Such side information can originate from previous transmissions or from other cooperating wiretappers which share some knowledge with each other.

In this paper we model the secrecy of the confidential message from a signal processing point of view. We require the wiretapper to have worst behavior of decoding performance regardless of the applied decoding strategy. For this secrecy criterion we establish the secrecy capacity and show that it equals the one of the classical wiretap channel (without side information). In addition, we derive necessary and sufficient conditions for a characterization of the corresponding optimal transceiver design.[1]

This work was partly supported by the German Ministry of Education and Research (BMBF) under Grant 01BQ1050.

2. WIRETAP CHANNEL

2.1. Classical Wiretap Channel without Side Information

In practical systems a transmitter usually uses a finite modulation scheme and a receiver quantizes the received signal before further processing so that it is reasonable to assume finite input and output alphabets denoted by $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. Then the channels $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and $V : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ represent the communication links to the legitimate receiver and the wiretapper respectively. For input and output sequences $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$, and $z^n \in \mathcal{Z}^n$ of block length n, the discrete memoryless channels are given by $W^n(y^n|x^n) := \prod_{i=1}^n W(y_i|x_i)$ and $V^n(z^n|x^n) := \prod_{i=1}^n V(z_i|x_i)$.

Definition 1. *An (n, J_n)-code $\mathcal{C}_n$ for the wiretap channel consists of a stochastic encoder at the transmitter*

$$E : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)$$

with a set of messages $\mathcal{J}_n := \{1, ..., J_n\}$ and a decoder at the legitimate receiver described by a collection of disjoint decoding sets

$$\{\mathcal{D}_j \subset \mathcal{Y}^n : j \in \mathcal{J}_n\}. \tag{1}$$

It is clear that every transmitter-receiver strategy results in a certain partition of the output alphabet as given in (1). This partition depends on the applied transmit and receive processing strategies.

Then for an (n, J_n)-code $\mathcal{C}_n$ the average and maximum probability of error are given by

$$\bar{e}(\mathcal{J}_n) := \frac{1}{|\mathcal{J}_n|} \sum_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(\mathcal{D}_j^c|x^n)$$

and

$$e_{\max}(\mathcal{J}_n) := \max_{j \in \mathcal{J}_n} \sum_{x^n \in \mathcal{X}^n} E(x^n|j) W^n(D_j^c|x^n).$$

To keep the transmitted message secret from the non-legitimate wiretapper, it is required

$$I(J; Z^n) \leq \epsilon_n \tag{2}$$

with J the random variable uniformly distributed over the set of messages $\mathcal{J}_n$ and $Z^n = (Z_1, Z_2, ..., Z_n)$ the output at the wiretapper. This criterion is known as *strong secrecy* [7, 8].

[1] *Notation:* $\mathcal{P}(\cdot)$ is the set of all probability distributions; $X - Y - Z$ denotes a Markov chain of random variables X, Y, and Z in this order; $\|\mu - \nu\|$ is the total variation distance of measures μ and ν.

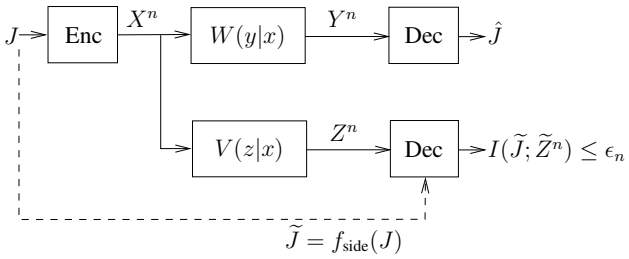

Fig. 1. Wiretap channel with side information. The side information $f_{\text{side}}(J)$ restricts the message to the subset $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$.

Definition 2. *A non-negative number R_S is an* achievable secrecy rate for the wiretap channel *if for all $\delta > 0$ there is an $n(\delta) \in \mathbb{N}$ and a sequence of (n, J_n)-codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that for all $n \geq n(\delta)$ we have $\frac{1}{n}\log J_n \geq R_S - \delta$ and*

$$I(J; Z^n) \leq \epsilon_n$$

while $\bar{e}(\mathcal{J}_n) \to 0$ (or $e_{max}(\mathcal{J}_n) \to 0$ respectively) and $\epsilon_n \to 0$ as $n \to \infty$. The secrecy capacity *C_S is given by the supremum of all achievable secrecy rates R_S.*

The classical wiretap channel is well studied for different secrecy criteria and its secrecy capacity can be found in [5–8, 14, 15].

Theorem 1. *The secrecy capacity C_S of the wiretap channel is*

$$C_S = \max_{V-X-(YZ)} \big(I(V;Y) - I(V;Z)\big).$$

2.2. Wiretap Channel with Side Information

In this paper the focus is on more powerful wiretappers. Additionally to its received channel output, the wiretapper has side information about the transmitted message available as depicted in Figure 1. Such side information can originate from prior transmissions due to a certain network structure or from other cooperating wiretappers which help each other to infer the confidential communication.

The side information at the wiretapper is modeled with the help of a deterministic function

$$f_{\text{side}} : \mathcal{J}_n \to \mathfrak{P}_2(\mathcal{J}_n)$$

with $\mathfrak{P}_2(\mathcal{J}_n)$ the power set of all subsets of $\mathcal{J}_n$ with cardinality at least 2. This means for transmitted message $J \in \mathcal{J}_n$, the wiretapper is aware of $f_{\text{side}}(J) \in \mathfrak{P}_2(\mathcal{J}_n)$ so that he (or she) can restrict the transmitted message to a subset $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$, i.e., he knows that the message belongs to $\widetilde{\mathcal{J}}$. The restriction $|\widetilde{\mathcal{J}}| \geq 2$ avoids the trivial case $|\widetilde{\mathcal{J}}| = 1$ where the transmitted message would be completely known to the wiretapper.

To incorporate the side information at the wiretapper, a natural extension of the security requirement (2) would be as follows. There has to be a universal ϵ_n (independent of the actual side information $\widetilde{\mathcal{J}}$) such that for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$ it holds

$$I(\widetilde{J}; \widetilde{Z}^n) \leq \epsilon_n \tag{3}$$

where $\widetilde{J}$ is the random variable uniformly distributed on the side information set $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ and $\widetilde{Z}^n = (\widetilde{Z}_1, \widetilde{Z}_2, ..., \widetilde{Z}_n)$ the corresponding output at the wiretapper with side information. With (3) the definition of an *achievable secrecy rate for the wiretap channel with side information* and the corresponding *secrecy capacity* follow accordingly as in Definition 2.

Instead of defining the secrecy by mutual information terms as in (3), here we use a criterion motivated from the signal processing point of view. We require worst behavior of decoding performance at the wiretapper regardless of the decoding strategy the wiretapper use. In more detail, for any side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$, the average probability of decoding error at the wiretapper has to satisfy

$$\bar{e}(\widetilde{\mathcal{J}}) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n \tag{4}$$

with $\lambda_n \to 0$ as $n \to \infty$. This means the decoding performance of the wiretapper is the same as if the wiretapper ignores its received signal and guesses the transmitted message based on its side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$. Thus, we require that the wiretapper does not take any advantage from its observation and simply selects a message $j \in \widetilde{\mathcal{J}}$ uniformly at random (regardless of its received $z^n \in \mathcal{Z}^n$). We call this a wiretapper with *maximum uncertainty*.

Definition 3. *A non-negative number R_S is an* achievable secrecy rate with maximum uncertainty for the wiretap channel with side information *if for all $\delta > 0$ there is an $n(\delta) \in \mathbb{N}$, a universal sequence $\{\lambda_n\}_{n\in\mathbb{N}}$, and a sequence of (n, J_n)-codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that for all $n \geq n(\delta)$ we have $\frac{1}{n}\log J_n \geq R_S - \delta$ and*

$$\bar{e}(\widetilde{\mathcal{J}}) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n$$

for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$, while $\bar{e}(\mathcal{J}_n) \to 0$ (or $e_{max}(\mathcal{J}_n) \to 0$ respectively) and $\lambda_n \to 0$ as $n \to \infty$. The secrecy capacity with maximum uncertainty *$C_{S,side}$ is given by the supremum of all achievable secrecy rates R_S with maximum uncertainty.*

For the analysis of the wiretap channel, it has been shown that the following property of a wiretap code is essential.

Definition 4. *A code for the wiretap channel (with side information) has exponentially fast* vanishing output variation at the wiretapper *if there is a measure ϑ on $\mathcal{Z}^n$ such that for all $j \in \mathcal{J}_n$ and*

$$\overline{V}^n(z^n|j) \coloneqq P_{Z|J}(z^n|j) = \sum_{x^n \in \mathcal{X}^n} E(x^n|j) V^n(z^n|x^n)$$

it holds

$$\big\|\overline{V}^n(\cdot|j) - \vartheta\big\| \leq \epsilon_n$$

with $\epsilon_n = 2^{-n\beta}$ for some $\beta > 0$.

In [14, 15] the vanishing output variation property was used to realize strong secrecy, cf. (2), for compound wiretap channels. Here, we show that it also allows realizing maximum uncertainty, cf. (4).

Proposition 1. *For any given code of Definition 1, the wiretapper with side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ has arbitrary decoding sets $\{\widetilde{\mathcal{D}}_j \subset \mathcal{Z}^n : j \in \widetilde{\mathcal{J}}\}$ with $\bigcup_{j\in\widetilde{\mathcal{J}}} \widetilde{\mathcal{D}}_j = \mathcal{Z}^n$. If the code has vanishing output variation, i.e., there is a measure ϑ on $\mathcal{Z}^n$ such that*

$$\big\|\overline{V}^n(\cdot|j) - \vartheta\big\| \leq 2^{-n\beta} \tag{5}$$

for all $j \in \mathcal{J}_n$, cf. Definition 4, then for the average probability of error $\bar{e}(\widetilde{\mathcal{J}})$ at the wiretapper it holds

$$\bar{e}(\widetilde{\mathcal{J}}) = \frac{1}{|\widetilde{\mathcal{J}}|} \sum_{j\in\widetilde{\mathcal{J}}} \overline{V}^n(\widetilde{\mathcal{D}}_j^c|j) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n \tag{6}$$

with $\lambda_n \to 0$ exponentially fast as $n \to \infty$.

Proof. If a code has the vanishing output variation property, i.e., it satisfies (5), then we have

$$\begin{aligned}\big\|P_{\widetilde{Z}^n\widetilde{J}} - P_{\widetilde{Z}^n}\otimes P_{\widetilde{J}}\big\| &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\big\|\overline{V}^n(\cdot|j) - P_{\widetilde{Z}^n}\big\| \\ &\le \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\Big(\big\|\overline{V}^n(\cdot|j)-\vartheta\big\| + \big\|\vartheta - P_{\widetilde{Z}^n}\big\|\Big) \\ &\le \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\Big(2^{-n\beta} + \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{k\in\widetilde{\mathcal{J}}}\big\|\vartheta - \overline{V}^n(\cdot|k)\big\|\Big) \\ &\le 2\cdot 2^{-n\beta} =: \lambda_n \end{aligned} \tag{7}$$

where $P_{\widetilde{Z}^n}\otimes P_{\widetilde{J}}$ is the product distribution defined by $(P_{\widetilde{Z}^n}\otimes P_{\widetilde{J}})(z^n,j) = P_{\widetilde{Z}^n}(z^n)P_{\widetilde{J}}(j)$ for all $z^n\in\mathcal{Z}^n$ and $j\in\widetilde{\mathcal{J}}$. The first inequality follows from the triangle inequality and the second and third inequalities from (5). On the other hand, we can write the average probability of error at the wiretapper as

$$\begin{aligned}\bar{e}(\widetilde{\mathcal{J}}) &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\overline{V}^n(\widetilde{\mathcal{D}}_j^c|j) = \sum_{j\in\widetilde{\mathcal{J}}}\overline{V}^n(\widetilde{\mathcal{D}}_j^c|j)P_{\widetilde{J}}(j) \\ &= \sum_{j\in\widetilde{\mathcal{J}}}P_{\widetilde{Z}^n\widetilde{J}}\big(\widetilde{\mathcal{D}}_j^c\times\{j\}\big) = P_{\widetilde{Z}^n\widetilde{J}}\Big(\bigcup_{j\in\widetilde{\mathcal{J}}}\widetilde{\mathcal{D}}_j^c\times\{j\}\Big). \end{aligned}\tag{8}$$

Now, with $\|P_{\widetilde{Z}^n\widetilde{J}} - P_{\widetilde{Z}^n}\otimes P_{\widetilde{J}}\|\le\lambda_n$ and $\lambda_n\to 0$ as $n\to\infty$, cf. (7), we can bound $\bar{e}(\widetilde{\mathcal{J}})$ in (8) from below by

$$\begin{aligned}\bar{e}(\widetilde{\mathcal{J}}) &\ge (P_{\widetilde{Z}^n}\otimes P_{\widetilde{J}})\Big(\bigcup_{j\in\widetilde{\mathcal{J}}}\widetilde{\mathcal{D}}_j^c\times\{j\}\Big) - \lambda_n \\ &= \sum_{j\in\widetilde{\mathcal{J}}}(P_{\widetilde{Z}^n}\otimes P_{\widetilde{J}})\big(\widetilde{\mathcal{D}}_j^c\times\{j\}\big) - \lambda_n = \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}P_{\widetilde{Z}^n}(\widetilde{\mathcal{D}}_j^c) - \lambda_n \\ &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\big(1 - P_{\widetilde{Z}^n}(\widetilde{\mathcal{D}}_j)\big) - \lambda_n = 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n \end{aligned}$$

where the last equality follows from the observation that for any non-negative numbers $a_1,...,a_N$ with $\sum_{i=1}^N a_i = 1$ we have $\sum_{i=1}^N(1-a_i) = N-1$. Note that λ_n is universal in the sense that it does not depend on the actual side information $\widetilde{\mathcal{J}}\subseteq\mathcal{J}_n$. □

Proposition 1 analyzed the pre-processing at the transmitter which leads to maximum uncertainty at the wiretapper. It established the property of vanishing output variation as a sufficient condition for maximum uncertainty and further showed that we can achieve maximum uncertainty exponentially fast.

3. SECRECY CAPACITY UNDER SIDE INFORMATION

In the following, we analyze the wiretap channel with side information in detail. In particular, we show that the secrecy capacity with maximum uncertainty of the wiretap channel with side information equals the secrecy capacity of the classical wiretap channel (without side information), cf. Theorem 1.

Theorem 2. *The secrecy capacity with maximum uncertainty of the wiretap channel with side information equals the secrecy capacity of the classical wiretap channel (without side information), i.e.,*

$$C_{S,side} = C_S.$$

3.1. Proof of Achievability

First, we prove the following inequality $C_{S,\text{side}}\ge C_S$ by giving an explicit construction of a transceiver design. Therefore, we need a wiretap code that realizes two tasks simultaneously: reliable communication at the desired rate C_S to the legitimate receiver, i.e., $e_{\max}(\mathcal{J}_n)\to 0$ as $n\to\infty$, and maximum uncertainty at the wiretapper, i.e., $\bar{e}(\widetilde{\mathcal{J}})\to 1-\frac{1}{|\widetilde{\mathcal{J}}|}$ for all $\widetilde{\mathcal{J}}\subseteq\mathcal{J}_n$ with $|\widetilde{\mathcal{J}}|\ge 2$ as $n\to\infty$.

To do so, we use a wiretap code with exponentially fast decreasing vanishing output variation, i.e., with $\epsilon_n = 2^{-n\beta}$ in Definition 4. This means the code has the property that there is a measure ϑ on $\mathcal{Z}^n$ such that for all $j\in\mathcal{J}_n$ we have

$$\big\|\overline{V}^n(\cdot|j) - \vartheta\big\| \le 2^{-n\beta}. \tag{9}$$

In [14, 15] it is shown that such a code (with vanishing output variation property) achieves the secrecy capacity of the wiretap channel (without side information), cf. Theorem 1. Thus, the first task, i.e., the reliable communication at the desired rate C_S is immediately given by [14, 15].

It remains to check, if the second task, i.e., the maximum uncertainty at the wiretapper with side information, is also satisfied. Since the code satisfies (9), we immediately obtain from Proposition 1 that the average decoding error at the wiretapper with side information satisfies $\bar{e}(\widetilde{\mathcal{J}})\ge 1-\frac{1}{|\widetilde{\mathcal{J}}|}-\lambda_n$ with $\lambda_n = 2\cdot 2^{-n\beta}$, $\beta>0$. Thus, the maximum uncertainty at the wiretapper is simultaneously guaranteed by the vanishing output variation property. This completes the proof of achievability.

3.2. Proof of Converse

The previous analysis has shown that for the wiretap channel with side information under the maximum uncertainty criterion (4), we can achieve the same rates as for the classical wiretap channel (without side information) under the strong secrecy criterion (2).

If we would analyze the wiretap channel with side information under the corresponding strong secrecy criterion (3), the inequality $C_{S,\text{side}}\le C_S$ would immediately follow, since additional side information at the wiretapper can only decrease the secrecy capacity. But here we consider secrecy based on the maximum uncertainty criterion (4), which relies on the decoding performance at the wiretapper and not on mutual information quantities. This makes the corresponding inequality by no means self-evident.

The following proposition allows to show that the inequality $C_{S,\text{side}}\le C_S$ holds also under the maximum uncertainty criterion. In addition, the result is interesting for itself as it characterizes the optimal pre-processing at the transmitter and establishes the vanishing output variation property also as a necessary condition for maximum uncertainty at the wiretapper.

Proposition 2. *Let $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be a sequence of wiretap codes achieving the secrecy capacity with maximum uncertainty of the wiretap channel with side information. Let $E:\mathcal{J}_n\to\mathcal{P}(\mathcal{X}^n)$ be the corresponding stochastic encoder and $\overline{V}^n(z^n|j) := \sum_{x^n\in\mathcal{X}^n}E(x^n|j)V^n(z^n|x^n)$. Then, there exists an $\epsilon_n = 2^{-n\beta}$, $\beta>0$ and a measure ϑ on $\mathcal{Z}^n$ such that for all $j\in\mathcal{J}_n$ it holds*

$$\big\|\overline{V}^n(\cdot|j) - \vartheta\big\| \le \epsilon_n,$$

i.e., the optimal code has the vanishing output variation property.

Proof. Let $\widetilde{\mathcal{J}} = \{j_1, j_2\}\subseteq\mathcal{J}_n$ be an arbitrary message subset with two elements. By the assumption of maximum uncertainty we

have at the wiretapper for arbitrary decoding sets $\widetilde{\mathcal{D}}_{j_1}$ and $\widetilde{\mathcal{D}}_{j_2}$ (with $\widetilde{\mathcal{D}}_{j_1} \cap \widetilde{\mathcal{D}}_{j_2} = \emptyset$ and $\widetilde{\mathcal{D}}_{j_1} \cup \widetilde{\mathcal{D}}_{j_2} = \mathcal{Z}^n$)

$$\bar{e}(\widetilde{\mathcal{J}}) = \frac{1}{|\widetilde{\mathcal{J}}|} \sum_{j \in \widetilde{\mathcal{J}}} \overline{V}^n(\widetilde{\mathcal{D}}_j^c|j) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n = \frac{1}{2} - \lambda_n \quad (10)$$

with universal $\lambda_n \to 0$ exponentially fast by assumption. Let $P_{\widetilde{Z}^n\widetilde{J}}(z^n, j) = \overline{V}^n(z^n|j)P_{\widetilde{J}}(j)$ be the joint distribution and $P_{\widetilde{J}}$ and $P_{\widetilde{Z}^n}$ be the marginals. Since the messages are uniformly distributed, we have $P_{\widetilde{J}}(j_1) = P_{\widetilde{J}}(j_2) = \frac{1}{2}$. We can write (10) as

$$\begin{aligned} &\frac{1}{2} \sum_{z^n \in \widetilde{\mathcal{D}}_{j_1}^c} \overline{V}^n(z^n|j_1) + \frac{1}{2} \sum_{z^n \in \widetilde{\mathcal{D}}_{j_2}^c} \overline{V}^n(z^n|j_2) \quad (11) \\ &\quad = \frac{1}{2}\Big(1 - \sum_{z^n \in \widetilde{\mathcal{D}}_{j_2}^c} \overline{V}^n(z^n|j_1) + \sum_{z^n \in \widetilde{\mathcal{D}}_{j_2}^c} \overline{V}^n(z^n|j_2)\Big) \geq \frac{1}{2} - \lambda_n \end{aligned}$$

where the equality follows from the substitutions $\widetilde{\mathcal{D}}_{j_1}^c = \widetilde{\mathcal{D}}_{j_2}$ and $\widetilde{\mathcal{D}}_{j_2} = \mathcal{Z}^n \backslash \widetilde{\mathcal{D}}_{j_2}^c$. This can easily be rewritten as

$$\sum_{z^n \in \widetilde{\mathcal{D}}_{j_2}^c} \big(\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big) \leq 2\lambda_n. \quad (12)$$

Since the decoding set $\widetilde{\mathcal{D}}_{j_2}$ can be arbitrary by assumption, we obtain for an arbitrary set $\mathcal{A} \subset \mathcal{Z}^n$ from (12) that $\sum_{z^n \in \mathcal{A}}(\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)) \leq 2\lambda_n$. Now, interchanging the roles of j_1 and j_2 and substituting $\widetilde{\mathcal{D}}_{j_2}^c = \widetilde{\mathcal{D}}_{j_1}$ in (11), we similarly obtain $\sum_{z^n \in \mathcal{A}}(\overline{V}^n(z^n|j_2) - \overline{V}^n(z^n|j_1)) \leq 2\lambda_n$ so that we end up with

$$\Big| \sum_{z^n \in \mathcal{A}} \overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\Big| \leq 2\lambda_n.$$

Let us define the sets

$$\begin{aligned} \mathcal{A}_+ &:= \big\{z^n \in \mathcal{Z}^n : \overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2) \geq 0\big\} \\ \mathcal{A}_- &:= \big\{z^n \in \mathcal{Z}^n : \overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2) < 0\big\} \end{aligned}$$

with $\mathcal{A}_- = (\mathcal{A}_+)^c$. Then

$$\begin{aligned} &\sum_{z^n \in \mathcal{A}_+} \big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big| \\ &\quad = \sum_{z^n \in \mathcal{A}_+} \big(\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big) \leq 2\lambda_n \quad (13) \end{aligned}$$

and similarly

$$\sum_{z^n \in \mathcal{A}_-} \big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big| \leq 2\lambda_n. \quad (14)$$

With $\mathcal{Z}^n = \mathcal{A}_+ \cup \mathcal{A}_-$ we conclude from (13) and (14) on

$$\begin{aligned} \big\|\overline{V}^n(\cdot|j_1) - \overline{V}^n(\cdot|j_2)\big\| &= \sum_{z^n \in \mathcal{Z}^n} \big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big| \\ &= \sum_{z^n \in \mathcal{A}_+} \big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big| \\ &\quad + \sum_{z^n \in \mathcal{A}_-} \big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big| \leq 4\lambda_n. \end{aligned}$$

Now, we set

$$\vartheta(z^n) = \frac{1}{|\mathcal{J}_n|} \sum_{j \in \mathcal{J}_n} \overline{V}^n(z^n|j)$$

for all $z^n \in \mathcal{Z}^n$, so that for any $k \in \widetilde{\mathcal{J}}$ we have

$$\begin{aligned} \big\|\overline{V}^n(\cdot|k) - \vartheta\big\| &= \Big\|\overline{V}^n(\cdot|k) - \frac{1}{|\widetilde{\mathcal{J}}|} \sum_{j \in \widetilde{\mathcal{J}}} \overline{V}^n(\cdot|j)\Big\| \\ &= \Big\|\frac{1}{|\widetilde{\mathcal{J}}|} \sum_{j \in \widetilde{\mathcal{J}}} \big(\overline{V}^n(\cdot|k) - \overline{V}^n(\cdot|j)\big)\Big\| \\ &\leq \frac{1}{|\widetilde{\mathcal{J}}|} \sum_{j \in \widetilde{\mathcal{J}}} \big\|\overline{V}^n(\cdot|k) - \overline{V}^n(\cdot|j)\big\| \leq 4\lambda_n =: \epsilon. \end{aligned}$$

This means an optimal code for the wiretap channel with side information and maximum uncertainty at the wiretapper always has to have the vanishing output variation property. □

Now, consider any code that achieves the secrecy capacity with maximum uncertainty $C_{S,\text{side}}$ of the wiretap channel with side information. From previous Proposition 2 follows that this code has the vanishing output variation property. From [14, 15] we know that if $\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}$ for all $j \in \mathcal{J}_n$, then $I(J; Z^n) \leq 2^{-n\frac{\beta}{2}}$ for n large enough. Thus, this code is also a good code for the wiretap channel (without side information) so that this code cannot achieve higher rates than C_S, cf. Theorem 1. This proves $C_{S,\text{side}} \leq C_S$.

4. DISCUSSION AND FURTHER EXTENSIONS

The wiretap channel under channel uncertainty is considered in [14, 15], where the compound wiretap channel for strong secrecy is studied. Here we assumed perfect channel state information at all users, but made extensively use of the code construction presented in [14, 15]. Having this in mind, it is straightforward to incorporate the effects of channel uncertainty by extending the results derived in this paper to the compound wiretap channel with side information.

If the channel to the wiretapper is not perfectly known, it can also be interpreted as multiple wiretappers, where each possible channel realization corresponds to another wiretapper. Thus, it captures also the scenario of multiple wiretappers so that our results can be extended to an optimal transceiver design that works universally for all wiretappers simultaneously. Due to space constraints we omit the details.

5. RELATION TO PRIOR WORK

The secrecy of transmitted messages is usually characterized using the (strong) secrecy criterion, cf. (2), which is based on mutual information terms, cf. for example [5–15]. In this paper, we use the concept of maximum uncertainty to characterize the secrecy of the confidential communication. This criterion is based on the decoding performance of non-legitimate receivers. Thus, it has an operational meaning in the sense that if maximum uncertainty is guaranteed, the wiretapper cannot decode the transmitted message regardless of the decoding strategy that is applied. This framework allows to explicitly characterize the optimal transceiver design which results in such a decoding performance at the wiretapper.

The available side information at the wiretapper is another point which distinguishes our paper from previous studies, cf. [5–15]. There, it is assumed that the wiretapper has only its received channel output available for decoding. Here, the wiretapper has additional side information available for further post-processing. Surprisingly, our results show that this side information does not lead to an decrease in secrecy capacity compared to the case of no side information.

6. REFERENCES

[1] Yingbin Liang, H. Vincent. Poor, and Shlomo Shamai (Shitz), "Information Theoretic Security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4-5, pp. 355–580, 2009.

[2] Eduard A. Jorswieck, Anne Wolf, and Sabrina Gerbracht, "Secrecy on the Physical Layer in Wireless Networks," *Trends in Telecommunications Technologies*, pp. 413–435, Mar. 2010.

[3] Ruoheng Liu and Wade Trappe, Eds., *Securing Wireless Communications at the Physical Layer*, Springer, 2010.

[4] Matthieu Bloch and João Barros, *Physical-Layer Security: From Information Theory to Security Engineering*, Cambridge University Press, 2011.

[5] Aaron D. Wyner, "The Wire-Tap Channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.

[6] Imre Csiszár and János Körner, "Broadcast Channels with Confidential Messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[7] Imre Csiszár, "Almost Independence and Secrecy Capacity," *Probl. Pered. Inform.*, vol. 32, no. 1, pp. 48–57, 1996.

[8] Ueli M. Maurer and Stefan Wolf, "Information-Theoretic Key Agreement: From Weak to Strong Secrecy for Free," in *EUROCRYPT 2000, Lecture Notes in Computer Science*, vol. 1807, pp. 351–368. Springer-Verlag, May 2000.

[9] S. K. Leung-Yan-Cheong and Martin E. Hellman, "The Gaussian Wire-Tap Channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, July 1978.

[10] Yingbin Liang, Gerhard Kramer, H. Vincent Poor, and Shlomo Shamai (Shitz), "Compound Wiretap Channels," *EURASIP J. Wireless Commun. Netw.*, vol. Article ID 142374, pp. 1–13, 2009.

[11] Ashish Khisti and Gregory W. Wornell, "Secure Transmission With Multiple Antennas I: The MISOME Wiretap Channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, July 2010.

[12] Ashish Khisti and Gregory W. Wornell, "Secure Transmission With Multiple Antennas–Part II: The MIMOME Wiretap Channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[13] Ersen Ekrem and Sennur Ulukus, "On Gaussian MIMO Compound Wiretap Channels," in *Proc. Conf. Inf. Sciences and Systems*, Baltimore, MD, USA, Mar. 2010, pp. 1–6.

[14] Igor Bjelaković, Holger Boche, and Jochen Sommerfeld, "Capacity Results for Compound Wiretap Channels," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 60–64.

[15] Igor Bjelaković, Holger Boche, and Jochen Sommerfeld, "Secrecy Results for Compound Wiretap Channels," *Probl. Inf. Transmission*, 2013, accepted.