Wiretap Channels With Side Information—Strong Secrecy Capacity and Optimal Transceiver Design

Holger Boche, *Fellow, IEEE*, and Rafael F. Schaefer, *Member, IEEE*

***Abstract*—The wiretap channel models the communication scenario where two legitimate users want to communicate in such a way that a wiretapper is kept ignorant. In this paper, the *wiretap channel with side information* is studied, where the wiretapper has additional side information available. This side information allows the wiretapper to restrict the transmitted message to a certain subset of messages before further postprocessing. Two different criteria are employed to model the secrecy of the confidential message: the information theoretic criterion of strong secrecy and a signal-processing-inspired criterion based on the decoding performance of the wiretapper. For the latter, the wiretapper is required to have the worst decoding performance regardless of the specific decoding strategy that is used. It is shown that both criteria are equivalent in terms of secrecy capacity. Furthermore, the secrecy capacity equals the one of the classical wiretap channel without side information available at the wiretapper. In addition, the corresponding capacity-achieving code structure and optimal transceiver design are characterized and properties are identified. Finally, extensions to channel uncertainty and multiple wiretappers are discussed.**

***Index Terms*—Decoding performance, optimal transceiver design, secrecy capacity, side information, strong secrecy, wiretap channel.**

I. Introduction

THE concept of physical layer, or information theoretic, security is becoming more and more attractive, since it solely uses the physical properties of a wireless channel to establish security. In this context, the concept of strong information theoretic secrecy is of particular interest, since this implies that the confidential information cannot be reproduced from the received signal regardless of the applied postprocessing at nonlegitimate receivers. 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 ignorant. The aim of the transmitter is to encode and transmit the confidential message in such a way that the legitimate receiver is able to decode the message and, at the same time, the wiretapper is prevented from inferring the confidential information from the received signal. In [5] Wyner studied the special case of degraded channels, which was then generalized to general discrete memoryless channels in [6] and to Gaussian channels in [7]. The secrecy capacity of the multiple-input multiple-output (MIMO) Gaussian wiretap channel is then independently derived in [8], [9] and [10]. Subsequently, the structure of the optimal transmit covariance matrix has been analyzed under the matrix power constraint in [11] and under the average power constraint in [12].

All these works have in common that the wiretapper has only the received channel output available for postprocessing. Here we study the *wiretap channel with side information*, where we consider a more powerful wiretapper having 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 due to certain network structures or from other cooperating wiretappers which share some knowledge with each other.

The classical wiretap channel is briefly reviewed in Section II. The wiretap channel with side information is then introduced in Section III. We model the secrecy of the transmitted message by two different criteria. The first criterion is based on (classical) information theoretic secrecy concepts and known as strong secrecy. The second one is motivated from a signal processing point of view, where the wiretapper is required to have the worst decoding performance regardless of the decoding strategy he (or she) applies. With worst decoding performance we refer to a wiretapper, which cannot extract any useful information from the received signal. Thus, the wiretapper only has his side information available and it remains for him to choose one message uniformly from this subset of messages. We show that both secrecy criteria—the information theoretic concept and the signal-processing-inspired concept—yield the same secrecy capacity. Moreover, it turns out that the secrecy capacity of the wiretap channel with side information equals the one of the classical wiretap channel (without side information). To this end, we derive necessary and sufficient conditions for a characterization of the corresponding optimal transceiver design. This is done in Sections IV and V under the assumption of perfect channel state information at all users.

After these considerations we move on to more realistic channel conditions by taking imperfect channel state information (CSI) into account. To do so, we assume that the legitimate

Manuscript received November 12, 2012; revised February 20, 2013 and June 16, 2013; accepted June 20, 2013. Date of publication June 27, 2013; date of current version July 17, 2013. This work was supported in part by the German Ministry of Education and Research (BMBF) under Grant 01BQ1050 and in part by the German Research Foundation (DFG) under Grant BO 1734/25-1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Carlo Blundo.

The authors are with the Lehrstuhl für Theoretische Informationstechnik, Technische Universität München, Munich 80290, Germany (e-mail: boche@tum.de; rafael.schaefer@tum.de).

Digital Object Identifier 10.1109/TIFS.2013.2271424

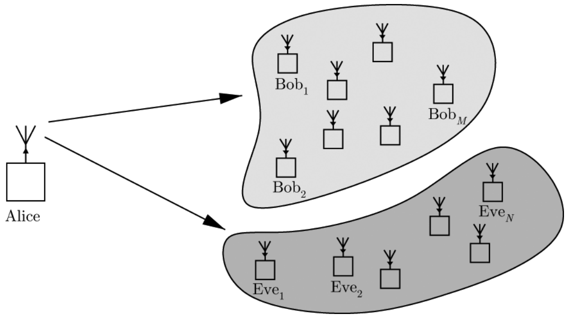


Fig. 1. Wiretap channel with multiple legitimate receivers and multiple wiretappers. The sender transmits a common (multicast) message to a group of M legitimate receivers (light gray) in the presence of a group of N wiretappers (dark gray) all to be kept ignorant of the transmitted message.

users do not know the exact channel realization which governs the transmission. Rather, it is only known that this channel realization belongs to a known set of channels and that it remains constant during the whole transmission of a codeword. This is the concept of *compound channels* [13], [14]. Besides channel uncertainty due to the nature of the wireless medium, it also captures implementation issues of practical systems such as imperfect or quantized channel estimation or limited feedback schemes. First studies for the (classical) compound wiretap channel can be found in [15]–[17]. In Section VI we study the corresponding *compound wiretap channel with side information* and extend results, previously obtained in Sections IV and V for the case of perfect CSI, to the case of compound channels. In addition, we discuss how these results can also be used to obtain results for the practically relevant case of multiple legitimate receivers and multiple wiretappers as visualized in Fig. 1. Finally, we conclude the paper in Section VII.

Notation

Discrete random variables are denoted by capital letters and their realizations and ranges by lower case and script letters, respectively; $\mathbb{N}$ and $\mathbb{R}_+$ are the sets of positive integers and nonnegative real numbers; $H(\cdot)$ and $I(\cdot;\cdot)$ are the traditional entropy and mutual information; $D(\cdot\|\cdot)$ is the Kullback-Leibler (information) divergence and $\|\mu-\nu\|$ is the total variation distance of measures μ and ν on $\mathcal{A}$ defined by $\|\mu-\nu\| := \sum_{a\in\mathcal{A}} |\mu(a)-\nu(a)|$; $X-Y-Z$ denotes a Markov chain of random variables X, Y, and Z in this order; all logarithms and information quantities are taken to the base 2; $\mathcal{P}(\cdot)$ is the set of all probability distributions; $(\mathcal{A})^c$, $|\mathcal{A}|$, and $\mathcal{A}\times\mathcal{B}$ are the complement, cardinality, and Cartesian product of the sets $\mathcal{A}$ and $\mathcal{B}$, respectively; the product distribution $P_A P_B$ is defined by the product marginal distributions of its components P_A and P_B, i.e., $P_A P_B(a,b) = P_A(a)P_B(b)$ for all elements a and b from their respective ranges $\mathcal{A}$ and $\mathcal{B}$; $\mathrm{lhs} := \mathrm{rhs}$ means the value of the right hand side (rhs) is assigned to the left hand side (lhs), $\mathrm{lhs} =: \mathrm{rhs}$ is defined accordingly.

II. Classical Wiretap Channel

In this section we start with the assumption of perfect channel state information (CSI) at all users and further assume that there is only one wiretapper to be kept ignorant. Later in Section VI, we will extend these results to the case of imperfect CSI and multiple wiretappers.

In practical systems, a transmitter usually uses a finite modulation scheme. For example, in a MIMO system, the signal to transmit is limited by a per-antenna power constraint so that finite modulation schemes, such as BPSK or QAM, are applied. Further, on the receiver side, the received signal is quantized before further processing. Thus, it is reasonable to assume finite input as well as finite output alphabets denoted by $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{Z}$ in the following. 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 $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, $z^n \in \mathcal{Z}^n$ 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 $V^n(z^n|x^n) := \prod_{i=1}^n V(z_i|x_i)$. Here, x_i and y_i, z_i are the input and output symbols at corresponding time instant i, $i = 1,\dots,n$. The wiretap channel is given by the pair of channels $\{W, V\}$ with common input.

In the classical wiretap channel, the task is to establish a reliable communication between the transmitter and the legitimate receiver and, at the same time, to keep the confidential information secret from the wiretapper. Most important, the wiretapper only has its channel output available to infer the confidential communication. This is formalized as follows.

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

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

i.e., a stochastic matrix, with a set of messages $\mathcal{J}_n$ and a decoder at the legitimate receiver described by a collection of disjoint decoding sets

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

Note that every transmitter-receiver strategy as specified above results in a certain partition of the output alphabet as $\bigcup_{j\in\mathcal{J}_n} \mathcal{D}_j = \mathcal{Y}^n$ with $\mathcal{D}_j \cap \mathcal{D}_k = \emptyset$ for $j \neq k$, cf. Equation (1). It is clear that the actual partition depends on the applied transmit and receive processing strategies.

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

$$\bar{e}_1(\mathcal{J}_n) := \frac{1}{|\mathcal{J}_n|} \sum_{j\in\mathcal{J}_n} \sum_{x^n\in\mathcal{X}^n} W^n(\mathcal{D}_j^c|x^n)E(x^n|j) \tag{2}$$

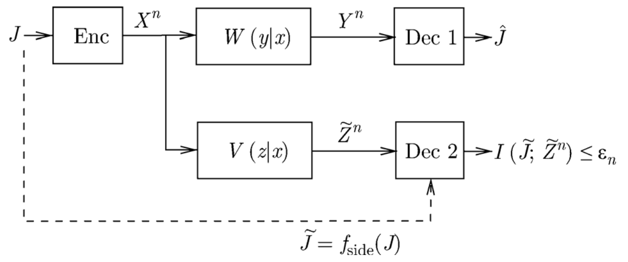
and

$$e_{1,\max}(\mathcal{J}_n) := \max_{j\in\mathcal{J}_n} \sum_{x^n\in\mathcal{X}^n} W^n(D_j^c|x^n)E(x^n|j). \tag{3}$$

To keep the transmitted message secret from the nonlegitimate wiretapper, the usual approach in the context of information theoretic security is to require

$$I(J;Z^n) \le \epsilon_n \tag{4}$$

with J the random variable uniformly distributed over the set of messages $\mathcal{J}_n$ and $Z^n = (Z_1, Z_2, \dots, Z_n)$ the channel output at



J
Enc
X^n
$W(y|x)$
Y^n
Dec 1
$\hat{J}$
$V(z|x)$
$\widetilde{Z}^n$
Dec 2
$I(\widetilde{J}; \widetilde{Z}^n) \leq \varepsilon_n$
$\widetilde{J} = f_{\text{side}}(J)$

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

the wiretapper. This criterion is known as *strong secrecy* [18], [19].

Definition 2: A nonnegative 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 $(n, \mathcal{J}_n)$-codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that for all $n \geq n(\delta)$ we have $1/n \log_2 |\mathcal{J}_n| \geq R_S - \delta$ and (4) is verified while $\bar{e}_1(\mathcal{J}_n) \to 0$ (or $e_{1,\max}(\mathcal{J}_n) \to 0$ 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 discrete memoryless wiretap channel is well studied under several aspects and its secrecy capacity can be found for instance in [5], [6], [18], [19].

Theorem 1: The secrecy capacity C_S of the wiretap channel is

$$C_S = \max_{U - X - (Y,Z)} \big(I(U;Y) - I(U;Z)\big)$$

where the random variables $U - X - (Y, Z)$ form a Markov chain.

Remark 1: The seminal works [5], [6] considered the weak secrecy criterion for the wiretap channel. More recently, the secrecy capacity of the wiretap channel has also been established for the strong secrecy criterion, cf. for example [17]–[19].

Remark 2: The secrecy capacity in Theorem 1 has been established for the average error criterion (2) as well as for the more stringent maximum error criterion (3), cf. also Definition 2. But from now on, we solely stick to the maximum error criterion for the analysis of the wiretap channel with side information.

III. Wiretap Channel With Side Information

In this paper the focus is on more powerful wiretappers. Additionally to his received channel output, the wiretapper has side information about the transmitted message available as depicted in Fig. 2. Such side information can be based on certain *a priori* knowledge at the wiretapper, but can also originate from prior transmissions due to a certain network structure or from other cooperating wiretappers which help each other inferring the confidential communication.

A. System Model

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 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 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.

The case $|\widetilde{\mathcal{J}}| = 2$ corresponds to the scenario with smallest uncertainty, since the wiretapper can narrow the transmitted message down to only two alternatives. In addition, the scenario of no side information available at the wiretapper is included in the model by the special case $\widetilde{\mathcal{J}} = \mathcal{J}_n$.

To model the secrecy of the transmitted message in the presence of side information available at the wiretapper, we study two alternative approaches: an information theory-based criterion and a signal-processing-inspired criterion.

B. Information Theoretic Security Approach

A natural extension of the information-theoretic-based approach of strong secrecy in (4) is the following. For any side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$, we require

$$I(\widetilde{J}; \widetilde{Z}^n) \leq \epsilon_n \tag{5}$$

where $\widetilde{J}$ is the random variable uniformly distributed on the side information set $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ and $\widetilde{Z}^n = (\widetilde{Z}_1, \widetilde{Z}_2, \ldots, \widetilde{Z}_n)$ the corresponding output at the wiretapper with side information.

Definition 3: A nonnegative number R_S is an *achievable secrecy rate for the wiretap channel with side information* if for all $\delta > 0$ there exists an $n(\delta) \in \mathbb{N}$, a universal sequence $\{\epsilon_n\}_{n\in\mathbb{N}}$ (in the sense that it is independent of the actual side information $\widetilde{\mathcal{J}}$), and a sequence of $(n, \mathcal{J}_n)$-codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that for all $n \geq n(\delta)$ we have $1/n \log_2 |\mathcal{J}_n| \geq R_S - \delta$ and (5) is verified for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$, while $e_{1,\max}(\mathcal{J}_n) \to 0$ and $\epsilon_n \to 0$ exponentially fast as $n \to \infty$. The *strong secrecy capacity* $C_{S,\text{side}}^{\text{strong}}$ is given by the supremum of all achievable secrecy rates R_S with strong secrecy.

Remark 3: In Definition 3 we explicitly require ϵ_n to decrease exponentially fast, i.e., to be of the form $\epsilon = 2^{-n\beta}$ for some $\beta > 0$. As we will see in the following, the applied methods actually provides an exponentially fast decrease so that this will be no restriction.

C. Signal Processing Approach

Instead of defining the secrecy by mutual information terms as done in (5), we also consider a criterion motivated from the signal processing point of view.

Similarly as for the legitimate receiver, cf. Definition 1, for any side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$, we can characterize every decoding strategy of the wiretapper by a collection of decoding sets

$$\{\widetilde{\mathcal{D}}_j \subset \mathcal{Z}^n : j \in \widetilde{\mathcal{J}}\}$$

with $\bigcup_{j\in\widetilde{\mathcal{J}}} \widetilde{\mathcal{D}}_j = \mathcal{Z}^n$ and $\widetilde{\mathcal{D}}_j \cap \widetilde{\mathcal{D}}_k = \emptyset$ for $j \neq k$. Note that the decoding sets at the wiretapper depend on the side information and, in general, are not the same for different side information. Then the average probability of decoding error at the wiretapper is

$$\bar{e}_2(\widetilde{\mathcal{J}}) = \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\sum_{x^n\in\mathcal{X}^n} V^n(\widetilde{\mathcal{D}}_j^c|x^n)E(x^n|j).$$

To ensure secrecy of the message, we require worst behavior of decoding performance at the wiretapper regardless of the decoding strategy that is used. In more detail, for any side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$, the average probability of decoding error has to satisfy

$$\bar{e}_2(\widetilde{\mathcal{J}}) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n \tag{6}$$

for some $\lambda_n > 0$ with $\lambda_n \to 0$ as $n \to \infty$ (we will specify λ_n later in Proposition 1). This means, for $n \to \infty$ the decoding performance of the wiretapper is the same as if the wiretapper ignores his received signal and guesses the transmitted message based on his side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$. Thus, we require that, asymptotically, the wiretapper does not take any advantage from his observation and simply selects a message $j \in \widetilde{\mathcal{J}}$ uniformly at random (regardless of his received $z^n \in \mathcal{Z}^n$). In this case, the probability of success is $1/|\widetilde{\mathcal{J}}|$, which is, in fact, the best the wiretapper can hope for. A wiretapper verifying (6) is said to be a wiretapper with *maximum uncertainty*.

Remark 4: We require (6) to hold for any decoding strategy the wiretapper may use. In particular, there are no restrictions imposed on the complexity of the decoding strategy. This will lead to universal results which hold for any postprocessing at the wiretapper.

Remark 5: The requirement of a high average decoding error at the wiretapper, cf. Equation (6), immediately implies that the maximum decoding error at the wiretapper is high as well.

Definition 4: A nonnegative 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 $(n, \mathcal{J}_n)$-codes $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ such that for all $n \geq n(\delta)$ we have $1/n \log_2 |\mathcal{J}_n| \geq R_S - \delta$ and (6) is verified for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$, while $e_{1,\max}(\mathcal{J}_n) \to 0$ and $\lambda_n \to 0$ exponentially fast as $n \to \infty$. The *secrecy capacity with maximum uncertainty* $C_{S,\text{side}}^{\text{uncert}}$ is given by the supremum of all achievable secrecy rates R_S with maximum uncertainty.

D. Vanishing Output Variation

In the following we want to analyze the secrecy capacity of the wiretap channel with side information under the strong secrecy and maximum uncertainty criterion in more detail. For this purpose, we extensively use the vanishing output variation property of a wiretap code, which has already been shown to be essential for the classical wiretap channel (without side information) [17].

Definition 5: A code for the wiretap channel (with side information) has exponentially fast *vanishing output variation at the wiretapper* if there exists a measure[1] ϑ on $\mathcal{Z}^n$ and some $\beta > 0$ such that for all $j \in \mathcal{J}_n$ and

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

it holds for all $z^n \in \mathcal{Z}^n$

$$\left\|\overline{V}^n(z^n|j) - \vartheta(z^n)\right\| \leq 2^{-n\beta}. \tag{7}$$

$$\begin{tikzcd}
\boxed{I(\widetilde{J};\widetilde{Z}^n) \leq \epsilon_n} \arrow[rr, dashed, "\text{Remark 6}"] \arrow[dr, shift right, "\text{Theorem 3}"'] & & \boxed{\bar{e}_2(\widetilde{\mathcal{J}}) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n} \arrow[dl, shift left, "\text{Theorem 5}"] \\
& \boxed{\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}} \arrow[ul, shift right, "\text{Proposition 1}"'] \arrow[ur, shift left, "\text{Proposition 1}"] &
\end{tikzcd}$$

Fig. 3. Relations between the strong secrecy criterion, the maximum uncertainty criterion, and the property of vanishing output variation. The arrows show the corresponding implications and indicate where it is proved.

In [17] the property of vanishing output variation was used to characterize the secrecy capacity of compound wiretap channels for the strong secrecy criterion, cf. Equation (4). Here, we show that it also allows realizing maximum uncertainty according to Definition 4, cf. also (6).

For clarity of presentation, Fig. 3 visualizes the relations between the previously introduced secrecy criteria and the vanishing output variation property for the analysis of the secrecy capacity of the wiretap channel with side information. The strong secrecy criterion (8) implies the maximum uncertainty criterion (6), which basically follows from Pinsker's inequality, see the following Remark 6. It is not obvious that the reverse implication also holds, since the strong secrecy criterion is in general a stronger assertion than the maximum uncertainty criterion.

In the following we show that the property of vanishing output variation allows establishing equivalence of both criteria in terms of secrecy capacity. We want to stress that this is only one possible strategy to realize equivalence of strong secrecy and maximum uncertainty in terms of secrecy capacity and there might be other approaches as well. However, the use of the vanishing output variation property is not only be suitable to prove the desired results, but also to highlight the practical relevance, since it characterizes the optimal preprocessing at the transmitter.

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 according to Definition 5, i.e., there is measure ϑ on $\mathcal{Z}^n$ such that $\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}$ for all $j \in \mathcal{J}_n$, cf. Equation (7), then the following two assertions hold:

i) The strong secrecy criterion satisfies

$$I(\widetilde{J};\widetilde{Z}^n) \leq \epsilon_n \tag{8}$$

[1] A measure ϑ on $\mathcal{Z}^n$ is assumed to satisfy the standard properties of nonnegativity, i.e., $\vartheta(\mathcal{A}) \geq 0$ for all $\mathcal{A} \subseteq \mathcal{Z}^n$, null empty set, i.e., $\vartheta(\emptyset) = 0$, and countable additivity, i.e., for all collections $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$ of pairwise disjoint sets it holds $\vartheta(\bigcup_{i\in\mathcal{I}} \mathcal{A}_i) = \sum_{i\in\mathcal{I}} \vartheta(\mathcal{A}_i)$. We do not require $\vartheta(\mathcal{Z}^n) = 1$, i.e., ϑ is not necessarily a probability measure.

with universal $\epsilon_n \longrightarrow 0$ exponentially fast as $n \longrightarrow \infty$

.

ii) The average probability of decoding error at the wiretapper satisfies

$$\bar{e}_2(\widetilde{\mathcal{J}}) \ge 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n \tag{9}$$

with universal $\lambda_n \longrightarrow 0$ exponentially fast as $n \longrightarrow \infty$.

Proof: We start with the proof of the first assertion. Let $P_{\widetilde{Z}^n\widetilde{\mathcal{J}}}(z^n, j) = \overline{V}^n(z^n|j)P_{\widetilde{\mathcal{J}}}(j)$ for all $z^n \in \mathcal{Z}^n$ and $j \in \widetilde{\mathcal{J}}$ be the joint distribution and $P_{\widetilde{Z}^n}$ and $P_{\widetilde{\mathcal{J}}}$ be the corresponding marginals where the latter is the uniform distribution over the restricted set $\widetilde{\mathcal{J}}$. If a code has the vanishing output variation property, i.e., it satisfies (7), then we have

$$\begin{aligned}\|P_{\widetilde{Z}^n\widetilde{\mathcal{J}}} - P_{\widetilde{Z}^n}P_{\widetilde{\mathcal{J}}}\| &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\|\overline{V}^n(\cdot|j) - P_{\widetilde{Z}^n}\|\\ &\le \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\Big(\|\overline{V}^n(\cdot|j) - \vartheta\| + \|\vartheta - P_{\widetilde{Z}^n}\|\Big)\\ &\le \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\Big(2^{-n\beta} + \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{k\in\widetilde{\mathcal{J}}}\|\vartheta - \overline{V}^n(\cdot|k)\|\Big)\\ &\le 2\cdot 2^{-n\beta} =: \lambda_n, \end{aligned}\tag{10}$$

where the first inequality follows from the triangle inequality and the second and third inequalities from (7).

Now, from the continuity of the entropy function, cf. for example [20, Lemma 1.2.7], we get

$$\begin{aligned}I(\widetilde{J}; \widetilde{Z}^n) &= H(\widetilde{Z}^n) + H(\widetilde{J}) - H(\widetilde{Z}^n, \widetilde{J})\\ &= H(P_{\widetilde{Z}^n}P_{\widetilde{J}}) - H(P_{\widetilde{Z}^n\widetilde{J}})\\ &\le -2\cdot 2^{-n\beta}\log_2(2\cdot 2^{-n\beta})\\ &\quad + 2n\cdot 2^{-n\beta}\log_2(|\widetilde{\mathcal{J}}||\mathcal{Z}|)\\ &\le 2^{-n\beta/2} =: \epsilon_n\end{aligned}$$

where the last inequality holds for n large enough which proves the first assertion (8). Note that the term $|\widetilde{\mathcal{J}}|$ can be bounded from above by $|\mathcal{J}_n|$ so that there is a universal ϵ_n independent of the actual side information $\widetilde{\mathcal{J}}$.

Next we move on to the proof of the second assertion (9). Therefore we write the average probability of decoding error at the wiretapper as

$$\begin{aligned}\bar{e}_2(\widetilde{\mathcal{J}}) &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\overline{V}^n(\widetilde{\mathcal{D}}_j^c|j)\\ &= \sum_{j\in\widetilde{\mathcal{J}}}\overline{V}^n(\widetilde{\mathcal{D}}_j^c|j)P_{\widetilde{J}}(j)\\ &= \sum_{j\in\widetilde{\mathcal{J}}}P_{\widetilde{Z}^n\widetilde{J}}\big(\widetilde{\mathcal{D}}_j^c\times\{j\}\big)\\ &= P_{\widetilde{Z}^n\widetilde{J}}\Big(\bigcup_{j\in\widetilde{\mathcal{J}}}\widetilde{\mathcal{D}}_j^c\times\{j\}\Big).\end{aligned}\tag{11}$$

Now, with $\|P_{\widetilde{Z}^n\widetilde{J}} - P_{\widetilde{Z}^n}P_{\widetilde{J}}\| \le \lambda_n$ and $\lambda_n \longrightarrow 0$ as $n \longrightarrow \infty$, cf. Equation (10), we can bound $\bar{e}_2(\widetilde{\mathcal{J}})$ in (11) from below by

$$\begin{aligned}\bar{e}_2(\widetilde{\mathcal{J}}) &\ge P_{\widetilde{Z}^n}P_{\widetilde{J}}\Big(\bigcup_{j\in\widetilde{\mathcal{J}}}\widetilde{\mathcal{D}}_j^c\times\{j\}\Big) - \lambda_n\\ &= \sum_{j\in\widetilde{\mathcal{J}}}P_{\widetilde{Z}^n}P_{\widetilde{J}}\big(\widetilde{\mathcal{D}}_j^c\times\{j\}\big) - \lambda_n\\ &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}P_{\widetilde{Z}^n}(\widetilde{\mathcal{D}}_j^c) - \lambda_n\\ &= \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\big(1 - P_{\widetilde{Z}^n}(\widetilde{\mathcal{D}}_j)\big) - \lambda_n\\ &= \frac{1}{|\widetilde{\mathcal{J}}|}\big(|\widetilde{\mathcal{J}}| - 1\big) - \lambda_n\\ &= 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n.\end{aligned}$$

Note that λ_n is universal in the sense that it does not depend on the actual side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$, cf. also (10). ■

Remark 6: From Proposition 1 follows that strong secrecy (8) immediately implies maximum uncertainty at the wiretapper with side information, cf. Equation (9). This can easily be seen by applying Pinsker's inequality, cf. for example [20, Problem 3.18], as

$$\begin{aligned}\epsilon_n &\ge I(\widetilde{J}; \widetilde{Z}^n) = D(P_{\widetilde{Z}^n\widetilde{J}}\|P_{\widetilde{Z}^n}P_{\widetilde{J}})\\ &\ge \frac{1}{\ln 2}\big\|P_{\widetilde{Z}^n\widetilde{J}} - P_{\widetilde{Z}^n}P_{\widetilde{J}}\big\|^2\end{aligned}$$

so that $\|P_{\widetilde{Z}^n\widetilde{J}} - P_{\widetilde{Z}^n}P_{\widetilde{J}}\| \le \lambda_n$ with $\lambda_n = \sqrt{2\ln 2\epsilon_n}$. With this, the desired implication follows immediately as from (10) onwards. This further gives strong secrecy the operational meaning in the sense that it implies worst behavior of decoding performance at the wiretapper according to(6).

Proposition 1 analyzed suitable preprocessing strategies for the transmitter to keep the message secret from the wiretapper. It established the property of vanishing output variation as a sufficient condition for strong secrecy and, in particular, maximum uncertainty. In addition, both can be achieved exponentially fast, cf. also Remark 3.

IV. Strong Secrecy Criterion

In this section we study the wiretap channel with side information for the information-theoretic-secrecy-based criterion of strong secrecy. To this end, we derive the corresponding strong secrecy capacity region and characterize the optimal (i.e., capacity-achieving) transceiver design.

A. Strong Secrecy Capacity

First, we analyze the strong secrecy capacity of the wiretap channel with side information and show that available side information at the wiretapper has no impact on the strong secrecy capacity.

Theorem 2: The strong secrecy capacity of the wiretap channel with side information equals the strong secrecy capacity of the wiretap channel without side information, i.e.,

$$C_{S,\text{side}}^{\text{strong}} = C_S.$$

Proof: To prove the desired result, we have to show the converse and achievability. The inequality $C_{S,\text{side}}^{\text{strong}} \le C_S$ is obviously true, since additional side information at the wiretapper cannot increase the corresponding secrecy capacity. This immediately establishes the converse and it only remains to prove that C_S is actually achievable also in the case of side information available at the wiretapper.

For the achievability we have to construct a wiretap code that realizes simultaneously reliable communication at the desired rate to the legitimate receiver, i.e., $e_{1,\max}(\mathcal{J}_n) \to 0$ as $n \to \infty$, and secrecy at the wiretapper, i.e., $I(\widetilde{J}; \widetilde{Z}^n) \to 0$ for all $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$ as $n \to \infty$.

In [17] it is shown by random coding arguments that there exist wiretap codes with exponentially fast decreasing vanishing output variation, cf. Definition 5, which achieve the strong secrecy capacity C_S of the wiretap channel (without side information). In particular, for the maximum probability of error we have $e_{1,\max}(\mathcal{J}_n) \to 0$ as $n \to \infty$ for all $R_S < C_S$, which immediately implies for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$ that $e_{1,\max}(\widetilde{\mathcal{J}}) \to 0$ holds as well. Note that [17] treats the compound wiretap channel, but clearly, this code construction also works for the noncompound case.

Remark 7: We want to emphasize that it is important to have for the original set $\mathcal{J}_n$ vanishing maximum probability of error $e_{1,\max}(\mathcal{J}_n) \to 0$ and not only vanishing average probability of error $\bar{e}_1(\mathcal{J}_n) \to 0$. Only this ensures that the maximum probability of error $e_{1,\max}(\widetilde{\mathcal{J}}) \to 0$ for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ vanishes as well. Furthermore, if we would require vanishing average probability of error for the subsets, i.e., $\bar{e}_1(\widetilde{\mathcal{J}}) = 1/|\widetilde{\mathcal{J}}| \sum_{j\in\widetilde{\mathcal{J}}} \sum_{x^n\in\mathcal{X}^n} W^n(\mathcal{D}_j^c|x^n)E(x^n|j) \to 0$ for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ (instead of the maximum error criterion), we also would end up with the need of vanishing maximum probability of error $e_{1,\max}(\mathcal{J}_n) \to 0$ for the original set $\mathcal{J}_n$. This is because vanishing average probability of error $\bar{e}_1(\mathcal{J}_n) \to 0$ for $\mathcal{J}_n$ does not imply $\bar{e}_1(\widetilde{\mathcal{J}}) \to 0$ for all $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$. Thus, the requirement of $e_{1,\max}(\mathcal{J}_n) \to 0$ is no restriction as it is necessary to control the average and maximum probability of error for all subsets as well.

Thus, we only have to check that this code also satisfies the strong secrecy requirement at the wiretapper with side information. Since the used code has the vanishing output variation property, cf. Definition 5, we know that there is a measure ϑ on $\mathcal{Z}^n$ such that for all $j \in \widetilde{\mathcal{J}}$ we have $\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}, \beta > 0$. Then Proposition 1 immediately implies that $I(\widetilde{J}; \widetilde{Z}^n) \leq \epsilon_n$ with ϵ_n independent of the actual side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ and $\epsilon_n \to 0$ exponentially fast as $n \to \infty$. ■

This shows that additional side information at the wiretapper does not have an impact on the strong secrecy capacity. Moreover, a code with vanishing output variation property originally developed for the wiretap channel without side information is also suitable to protect the transmitted message in the event of additional side information at the wiretapper.

B. Optimal Preprocessing

The previous analysis showed that a wiretap code with vanishing output variation is sufficient to achieve the strong secrecy capacity of the wiretap channel with side information. Here we want to show that the reverse statement is also true. In more detail, the following result shows that an optimal code for the wiretap channel with side information necessarily has to have the vanishing output variation property.

Theorem 3: Let $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be a sequence of wiretap codes achieving the strong secrecy capacity of the wiretap channel with side information and $E : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)$ be the corresponding stochastic encoder. Then, there exists a measure ϑ on $\mathcal{Z}^n$ and some $\beta > 0$ such that for all $j \in \mathcal{J}_n$ it holds $\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}$, cf. Equation (7). This means, the optimal code has the vanishing output variation property according to Definition 5.

Proof: Let $\widetilde{\mathcal{J}} = \{j_1, j_2\} \subseteq \mathcal{J}_n$ be an arbitrary message subset with two elements. Since the code is optimal for the wiretap channel with side information by assumption, we have $I(\widetilde{J}; \widetilde{Z}^n) \leq \epsilon_n$ with $\epsilon_n \to 0$ exponentially fast as $n \to \infty$ and ϵ_n independent of $\widetilde{\mathcal{J}}$, cf. Definition 3 and (5). Let $P_{\widetilde{Z}^n\widetilde{J}}(z^n, j) = \overline{V}^n(z^n|j)P_{\widetilde{J}}(j)$ for all $z^n \in \mathcal{Z}^n$ and $j \in \widetilde{\mathcal{J}}$ be the joint distribution and $P_{\widetilde{Z}^n}$ and $P_{\widetilde{J}}$ be the corresponding marginals.

Writing (5) in terms of Kullback-Leibler (information) divergence as

$$I(\widetilde{J}; \widetilde{Z}^n) = D(P_{\widetilde{Z}^n\widetilde{J}} \| P_{\widetilde{Z}^n}P_{\widetilde{J}}) \leq \epsilon_n,$$

we get by Pinsker's inequality, cf. for example [20, Problem 3.18], for some constant $c > 0$ the following

$$\begin{aligned} c\sqrt{\epsilon_n} &\geq \left\|P_{\widetilde{Z}^n\widetilde{J}} - P_{\widetilde{Z}^n}P_{\widetilde{J}}\right\| \\ &= \sum_{z^n\in\mathcal{Z}^n} \left|P_{\widetilde{Z}^n\widetilde{J}}(z^n, j_1) - P_{\widetilde{Z}^n}(z^n)P_{\widetilde{J}}(j_1)\right| \\ &\quad + \sum_{z^n\in\mathcal{Z}^n} \left|P_{\widetilde{Z}^n\widetilde{J}}(z^n, j_2) - P_{\widetilde{Z}^n}(z^n)P_{\widetilde{J}}(j_2)\right| \\ &= \sum_{z^n\in\mathcal{Z}^n} \left|\frac{1}{2}\overline{V}^n(z^n|j_1) - \frac{1}{2}P_{\widetilde{Z}^n}(z^n)\right| \\ &\quad + \sum_{z^n\in\mathcal{Z}^n} \left|\frac{1}{2}\overline{V}^n(z^n|j_2) - \frac{1}{2}P_{\widetilde{Z}^n}(z^n)\right|. \end{aligned}$$

Here, the first equality follows from the definition of total variation distance and the fact that $\widetilde{\mathcal{J}}$ has only two elements, and the second equality from the fact that $\widetilde{J}$ is uniformly distributed on $\widetilde{\mathcal{J}}$. Thus, for each $l = 1, 2$ we have

$$\sum_{z^n\in\mathcal{Z}^n} \left|\overline{V}^n(z^n|j_l) - P_{\widetilde{Z}^n}(z^n)\right| \leq 2c\sqrt{\epsilon_n}.$$

Since

$$P_{\widetilde{Z}^n}(z^n) = \frac{1}{2}\left(\overline{V}^n(z^n|j_1) + \overline{V}^n(z^n|j_2)\right),$$

we have

$$\begin{aligned} &\sum_{z^n\in\mathcal{Z}^n} \left|\overline{V}^n(z^n|j_1) - \frac{1}{2}\overline{V}^n(z^n|j_1) - \frac{1}{2}\overline{V}^n(z^n|j_2)\right| \\ &\qquad = \frac{1}{2}\sum_{z^n\in\mathcal{Z}^n} \left|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\right| \leq 2c\sqrt{\epsilon_n}. \end{aligned}$$

Thus, by the definition of total variation distance we have for arbitrary $j_1, j_2 \in \mathcal{J}_n$

$$\begin{aligned} &\sum_{z^n\in\mathcal{Z}^n} \left|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\right| \\ &\qquad = \left\|\overline{V}^n(\cdot|j_1) - \overline{V}^n(\cdot|j_2)\right\| \leq 4c\sqrt{\epsilon_n}. \end{aligned}$$

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 $l \in \mathcal{J}_n$ we have

$$\begin{aligned} \|\overline{V}^n(\cdot|l) - \vartheta\| &= \left\|\overline{V}^n(\cdot|l) - \frac{1}{|\mathcal{J}_n|}\sum_{j\in\mathcal{J}_n}\overline{V}^n(\cdot|j)\right\| \\ &= \left\|\frac{1}{|\mathcal{J}_n|}\sum_{j\in\mathcal{J}_n}\left(\overline{V}^n(\cdot|l) - \overline{V}^n(\cdot|j)\right)\right\| \\ &\leq \frac{1}{|\mathcal{J}_n|}\sum_{j\in\mathcal{J}_n}\|\overline{V}^n(\cdot|l) - \overline{V}^n(\cdot|j)\| \\ &\leq 4c\sqrt{\epsilon_n}. \end{aligned} \quad (12)$$

This means an optimal code for the wiretap channel with side information always has the vanishing output variation property. Finally, since ϵ_n decreases exponentially fast by assumption, it is ensured that (12) is of the form $\|\overline{V}^n(\cdot|l) - \vartheta\| \leq 2^{-n\beta}$ as desired, cf. Equation (7). This completes the proof of the theorem. ■

From Proposition 1 we already know that the vanishing output variation property is a sufficient condition for realizing strong secrecy. Now, Theorem 3 establishes this also as a necessary condition by showing that an optimal code necessarily has to have this property. This characterizes the optimal preprocessing of the transmitter to establish strong secrecy in wiretap channels with side information.

V. Maximum Uncertainty Criterion

In this section we study the wiretap channel with side information for the signal-processing-inspired criterion of maximum uncertainty. To this end, we derive the corresponding secrecy capacity with maximum uncertainty and characterize the optimal (i.e., capacity-achieving) transceiver design.

A. Secrecy Capacity With Maximum Uncertainty

Theorem 4: The secrecy capacity with maximum uncertainty of the wiretap channel with side information equals the corresponding strong secrecy capacity of the classical wiretap channel without side information (and therewith also equals the strong secrecy capacity of the wiretap channel with side information), i.e.,

$$C_{S,\mathrm{side}}^{\mathrm{uncert}} = C_S.$$

Proof: We start with the achievability part and prove the following inequality $C_{S,\mathrm{side}}^{\mathrm{uncert}} \geq C_S$ by giving an explicit construction of a transceiver design. To do so, we need a wiretap code that realizes two tasks simultaneously: reliable communication at the desired rate to the legitimate receiver, i.e., $e_{1,\max}(\mathcal{J}_n) \to 0$ as $n \to \infty$, and maximum uncertainty at the wiretapper, i.e., $\bar{e}_2(\widetilde{\mathcal{J}}) \to 1 - 1/|\widetilde{\mathcal{J}}|$ for all $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$ as $n \to \infty$.

Similarly as in the proof of Theorem 2, we use a wiretap code with exponentially fast decreasing vanishing output variation, cf. Definition 5, which achieves the strong secrecy capacity C_S of the wiretap channel (without side information), cf. [17]. In particular, for the maximum probability of error we have $e_{1,\max}(\mathcal{J}_n) \to 0$ as $n \to \infty$ for all $R_S < C_S$, which further implies for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$ that $e_{1,\max}(\widetilde{\mathcal{J}}) \to 0$ holds as well.

Thus, 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 has the vanishing output variation property, cf. Definition 5, we immediately obtain from the second assertion of Proposition 1 that the average decoding error at the wiretapper with side information satisfies $\bar{e}_2(\widetilde{\mathcal{J}}) \geq 1 - 1/|\widetilde{\mathcal{J}}| - \lambda_n$ with $\lambda_n \to 0$ exponentially fast as $n \to \infty$, cf. Equation (9). Thus, the maximum uncertainty at the wiretapper is simultaneously guaranteed by the vanishing output variation property. This completes the proof of achievability.

It remains to prove the converse. If we would analyze the wiretap channel with side information under the corresponding strong secrecy criterion (5), the inequality $C_{S,\mathrm{side}}^{\mathrm{uncert}} \leq C_S$ would immediately follow as in Theorem 2, since additional side information at the wiretapper cannot increase the secrecy capacity. But here we consider secrecy based on the maximum uncertainty criterion (6), 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 and is therefore devoted to the next subsection.

B. Optimal Preprocessing

The previous analysis has shown that for the wiretap channel with side information under the maximum uncertainty criterion (6), we can achieve the same rates as for classical wiretap channel (without side information) under the strong secrecy criterion (4).

The following theorem allows to show that the inequality $C_{S,\mathrm{side}}^{\mathrm{uncert}} \leq C_S$ holds also under the maximum uncertainty criterion. In addition, the result is interesting in itself as it characterizes the optimal preprocessing at the transmitter and establishes the vanishing output variation property also as a necessary condition for maximum uncertainty at the wiretapper.

Theorem 5: 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 and $E : \mathcal{J}_n \to \mathcal{P}(\mathcal{X}^n)$ be the corresponding stochastic encoder. Then, there exists a measure ϑ on $\mathcal{Z}^n$ and some $\beta > 0$ such that for all $j \in \mathcal{J}_n$ it holds $\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}$, cf. Equation (7). This means the optimal code has the vanishing output variation property according to Definition 5.

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 $\widetilde{\mathcal{D}}_{j_1} \cup \widetilde{\mathcal{D}}_{j_2} = \mathcal{Z}^n$)

$$\bar{e}_2(\widetilde{\mathcal{J}}) = \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}} \overline{V}^n(\widetilde{\mathcal{D}}_j^c|j) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n = \frac{1}{2} - \lambda_n \quad (13)$$

with universal $\lambda_n \to 0$ exponentially fast by assumption. Let $P_{\widetilde{Z}^n\widetilde{J}}(z^n,j) = \overline{V}^n(z^n|j)P_{\widetilde{J}}(j)$ for all $z^n \in \mathcal{Z}^n$ and $j \in \widetilde{\mathcal{J}}$ be the joint distribution and $P_{\widetilde{J}}$ and $P_{\widetilde{Z}^n}$ be the corresponding marginals. Since the messages are uniformly distributed, we have $P_{\widetilde{J}}(j_1) = P_{\widetilde{J}}(j_2) = 1/2$. We can write (13) as

$$\begin{aligned}\frac{1}{2}\sum_{z^n\in\widetilde{\mathcal{D}}_{j_1}^c} &\overline{V}^n(z^n|j_1) + \frac{1}{2}\sum_{z^n\in\widetilde{\mathcal{D}}_{j_2}^c} \overline{V}^n(z^n|j_2)\\ &= \frac{1}{2}\Big(1 - \sum_{z^n\in\widetilde{\mathcal{D}}_{j_2}^c}\overline{V}^n(z^n|j_1) + \sum_{z^n\in\widetilde{\mathcal{D}}_{j_2}^c}\overline{V}^n(z^n|j_2)\Big)\\ &\geq \frac{1}{2} - \lambda_n \end{aligned}\tag{14}$$

where the equality follows from the substitutions $\widetilde{\mathcal{D}}_{j_1}^c = \widetilde{\mathcal{D}}_{j_2}$ and $\widetilde{\mathcal{D}}_{j_2} = \mathcal{Z}^n\backslash\widetilde{\mathcal{D}}_{j_2}^c$ (recall that there are only two disjoint decoding sets). This can easily be rewritten as

$$\sum_{z^n\in\widetilde{\mathcal{D}}_{j_2}^c}\big(\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big) \leq 2\lambda_n. \tag{15}$$

Since the decoding set $\widetilde{\mathcal{D}}_{j_2}$ can be arbitrary by assumption, we obtain from (15) for an arbitrary set $\mathcal{A}\subset\mathcal{Z}^n$ that

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

Now, interchanging the roles of j_1 and j_2 and substituting $\widetilde{\mathcal{D}}_{j_2}^c = \widetilde{\mathcal{D}}_{j_1}$ in (14), we similarly obtain

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

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 $\mathcal{A}_- = (\mathcal{A}_+)^c$ and $\mathcal{A}_+\cup\mathcal{A}_- = \mathcal{Z}^n$. Then

$$\begin{aligned}\sum_{z^n\in\mathcal{A}_+}&\big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big|\\ &= \sum_{z^n\in\mathcal{A}_+}\big(\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big) \leq 2\lambda_n\end{aligned}\tag{16}$$

and similarly

$$\begin{aligned}\sum_{z^n\in\mathcal{A}_-}&\big|\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big|\\ &= \sum_{z^n\in\mathcal{A}_-}\big(-\big(\overline{V}^n(z^n|j_1) - \overline{V}^n(z^n|j_2)\big)\big)\\ &= \sum_{z^n\in\mathcal{A}_-}\big(\overline{V}^n(z^n|j_2) - \overline{V}^n(z^n|j_1)\big) \leq 2\lambda_n.\end{aligned}\tag{17}$$

With $\mathcal{Z}^n = \mathcal{A}_+\cup\mathcal{A}_-$ we conclude from (16) and (17) 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 $l\in\widetilde{\mathcal{J}}$ we have

$$\begin{aligned}\big\|\overline{V}^n(\cdot|l) - \vartheta\big\| &= \Big\|\overline{V}^n(\cdot|l) - \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\overline{V}^n(\cdot|j)\Big\|\\ &= \Big\|\frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\big(\overline{V}^n(\cdot|l) - \overline{V}^n(\cdot|j)\big)\Big\|\\ &\leq \frac{1}{|\widetilde{\mathcal{J}}|}\sum_{j\in\widetilde{\mathcal{J}}}\big\|\overline{V}^n(\cdot|l) - \overline{V}^n(\cdot|j)\big\|\\ &\leq 4\lambda_n.\end{aligned}\tag{18}$$

This means an optimal code for the wiretap channel with side information and maximum uncertainty at the wiretapper always has the vanishing output variation property. Finally, since λ_n decreases exponentially fast by assumption, it is ensured that (18) is of the form $\|\overline{V}^n(\cdot|l) - \vartheta\| \leq 2^{-n\beta}$ as desired, cf. Equation (7). This completes the proof of the theorem. ■

Now we go back to the proof of converse of Theorem 4. Therefore, we consider any code that achieves the secrecy capacity with maximum uncertainty $C_{S,\mathrm{side}}^{\mathrm{uncert}}$ of the wiretap channel with side information. Now, from Theorem 5 we know that this code has the vanishing output variation property. From Proposition 1 we know that if $\|\overline{V}^n(\cdot|j) - \vartheta\| \leq 2^{-n\beta}$ for all $j\in\mathcal{J}_n$, then $I(\widetilde{J};\widetilde{Z}^n) \leq \epsilon$ for all $\widetilde{\mathcal{J}}\subseteq\mathcal{J}_n$. Thus, this code is also a good code for the wiretap channel with side information under the strong secrecy criterion so that this code cannot achieve higher rates than $C_{S,\mathrm{side}}^{\mathrm{strong}}$, cf. Theorem 2. This finally proves $C_{S,\mathrm{side}}^{\mathrm{uncert}} \leq C_{S,\mathrm{side}}^{\mathrm{strong}} = C_S$ and completes the proof of Theorem 4. ■

VI. Extensions to Channel Uncertainty and Multiple Wiretappers

In the previous analysis we assumed that transmitter and receiver both have perfect channel state information (CSI) and that there is only one wiretapper to be kept secret. In this section we extend these results to more realistic communication scenarios by taking channel uncertainty at the legitimate users into account. In contrast to this, we allow the wiretapper to have perfect CSI, but it is shown that the wiretapper will not take any advantage from this knowledge. Further, we discuss how these results yield also robust processing strategies to keep multiple wiretappers ignorant of the confidential communication. The aim of this section is to demonstrate the universality of the derived techniques in the sense that they immediately apply to much more involved problems as well.

A. Compound Wiretap Channel

Robust processing strategies are desirable in practical systems as there is always uncertainty in the channel state information due to the nature of the wireless medium but also due to implementational issues such as imperfect channel estimation or limited feedback schemes. A reasonable model is to assume that the exact channel realization is not known; rather, it is only known that it belongs to a prespecified set of channels. If this channel remains fixed during the whole transmission of a codeword, this corresponds to the concept of *compound channels* [13], [14]. This model captures the nature of the wireless medium, but also includes implementational issues of practical systems.

To model the compound wiretap channel, let $\mathcal{T}$ be a finite index set. Then for fixed $t \in \mathcal{T}$, the discrete memoryless channels to the legitimate receiver and the wiretapper are given by $W_t^n(y^n|x^n) := \prod_{i=1}^n W_t(y_i|x_i)$ and $V_t^n(z^n|x^n) := \prod_{i=1}^n V_t(z_i|x_i)$.

Definition 6: The discrete memoryless *compound wiretap channel* $\mathfrak{W}$ is given by

$$\mathfrak{W} := \big\{(W_t, V_t) : t \in \mathcal{T}\big\}.$$

Remark 8: This includes the widely adopted model of the form $\mathfrak{W} = \{(W_s, V_t) : s \in \mathcal{S}, t \in \mathcal{T}\}$ with $\mathcal{S} \neq \mathcal{T}$ as we can always construct a new set of the form $\hat{\mathcal{T}} = \mathcal{S} \times \mathcal{T}$.

Since transmitter and receiver do not know the exact channel realization, they have to use universal encoder and decoder not depending on the actual channel realization similarly as already given in Definition 1. The only difference is that we have to ensure that the legitimate receiver can decode the transmitted message for all channel realizations $t \in \mathcal{T}$ so that the average probability and maximum probability of error become

$$\bar{e}_1(\mathcal{J}_n) := \max_{t\in\mathcal{T}} \frac{1}{|\mathcal{J}_n|} \sum_{j\in\mathcal{J}_n} \sum_{x^n\in\mathcal{X}^n} W_t^n(\mathcal{D}_j^c|x^n)E(x^n|j)$$

and

$$e_{1,\max}(\mathcal{J}_n) := \max_{t\in\mathcal{T}} \max_{j\in\mathcal{J}_n} \sum_{x^n\in\mathcal{X}^n} W_t^n(\mathcal{D}_j^c|x^n)E(x^n|j).$$

Accordingly, since the exact channel realization to the wiretapper is not known at the legitimate users, we have to ensure that the message is kept secret for all possible channel realizations $t \in \mathcal{T}$. Thus, the strong secrecy requirement in (5) reads now as

$$\max_{t\in\mathcal{T}} I(\widetilde{J}; \widetilde{Z}_t^n) \leq \epsilon_n \tag{19}$$

for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$ and $\widetilde{Z}_t^n = (\widetilde{Z}_{t,1}, \widetilde{Z}_{t,2}, \ldots, \widetilde{Z}_{t,n})$ the corresponding output at the wiretapper for channel realization $t \in \mathcal{T}$. On the other hand, the maximum uncertainty criterion in (6) is independent of the actual channel realization. Thus, it remains the same also in the compound setting, i.e.,

$$\bar{e}_2(\widetilde{\mathcal{J}}) \geq 1 - \frac{1}{|\widetilde{\mathcal{J}}|} - \lambda_n \tag{20}$$

for all subsets $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}| \geq 2$.

With these extensions the definition of an *achievable secrecy rate for the compound wiretap channel with side information* and the corresponding *strong secrecy capacity* $C_{S,\text{side}}^{\text{strong}}(\mathfrak{W})$ and *secrecy capacity with maximum uncertainty* $C_{S,\text{side}}^{\text{uncert}}(\mathfrak{W})$ follow accordingly from Definitions 3 and 4.

The previous analysis for perfect CSI has shown that the property of vanishing output variation at the wiretapper, cf. Definition 5, plays an important role. And so it does for the compound wiretap channel so that we slightly adapt it as stated in [17].

Definition 7: A code for the compound wiretap channel (with side information) has exponentially fast *vanishing output variation* if for all $t \in \mathcal{T}$ there are measures ϑ_t on $\mathcal{Z}^n$ and some $\beta > 0$ such that for all $j \in \mathcal{J}_n$ and

$$\overline{V}_t^n(z^n|j) := \sum_{x^n\in\mathcal{X}^n} V_t^n(z^n|x^n)E(x^n|j)$$

it holds for all $z^n \in \mathcal{Z}^n$

$$\big\|\overline{V}_t^n(z^n|j) - \vartheta_t(z^n)\big\| \leq 2^{-n\beta}.$$

The difference to Definition 5 is that it is sufficient to have for each channel realization $t \in \mathcal{T}$ a different measure ϑ_t which need not be the same for different channel realizations. This is can be justified by the observation that for a certain channel realization we have to ensure that the output at the wiretapper looks "similar" for all possible messages. But it is not important that it is the same for all different channel realizations. With this and [17] we get the following.

Theorem 6: An achievable secrecy rate for the compound wiretap channel with side information (for both strong secrecy and maximum uncertainty) is given by

$$R_S = \max_{U-X-(Y_t,Z_t)} \big(\min_{t\in\mathcal{T}} I(U;Y_t) - \max_{t\in\mathcal{T}} I(U;Z_t)\big) \tag{21}$$

where the random variables $U - X - (Y_t, Z_t)$ form a Markov chain, where Y_t and Z_t are the corresponding output random variables for channel realization $t \in \mathcal{T}$.

Sketch of Proof: Basically, the achievability of the rate given in (21) follows the one in Theorem 2, where we again make use of the code construction given in [17]. In more detail, it already presents the construction of a code for the wiretap channel (without side information) which establishes a reliable communication to the legitimate receiver at the desired rate.

Thus, as in the proof of Theorem 2, it remains to check, if the corresponding security requirements (19) and (20) are fulfilled. Since the code from [17] has the vanishing output variation property, cf. Definition 7, it can easily be verified that for each $t \in \mathcal{T}$ the property $\|\overline{V}_t^n(\cdot|j) - \vartheta_t\| \leq 2^{-n\beta}$ implies $I(\widetilde{J}; \widetilde{Z}_t^n) \leq \epsilon_n$ as well as $\bar{e}_2(\widetilde{\mathcal{J}}) \geq 1 - 1/|\widetilde{\mathcal{J}}| - \lambda_n$, cf. Proposition 1. We omit the details for brevity. ■

For the following discussion, we stick to the strong secrecy criterion (19). Since additional side information at the wiretapper can only decrease the secrecy capacity, every upper bound on the strong secrecy capacity of the compound wiretap channel (without side information) immediately yields an upper bound on the corresponding capacity for the case with side information at the wiretapper. Accordingly, from [17] we get the following multiletter upper bound.

Theorem 7: An upper bound on the strong secrecy capacity of the compound wiretap channel with side information is given by

$$C_{S,\text{side}}^{\text{strong}}(\mathfrak{W}) \leq \lim_{n \longrightarrow \infty} \frac{1}{n} \max_{U - X^n - (Y_t^n, Z_t^n)} \times \big(\min_{t \in \mathcal{T}} I(U; Y_t^n) - \max_{t \in \mathcal{T}} I(U; Z_t^n) \big)$$

where the random variables $U - X^n - (Y_t^n, Z_t^n)$ form a Markov chain.

Applying the achievability result given in Theorem 7 to the n-th memoryless extension of the channels, i.e., (W_t^n, V_t^n), yields the corresponding multiletter case. Together with the converse result given in Theorem 7 we conclude on the following.

Corollary 1: A multiletter description of the strong secrecy capacity of the compound wiretap channel with side information is given by

$$C_{S,\text{side}}^{\text{strong}}(\mathfrak{W}) = \lim_{n \longrightarrow \infty} \frac{1}{n} \max_{U - X^n - (Y_t^n, Z_t^n)} \times \big(\min_{t \in \mathcal{T}} I(U; Y_t^n) - \max_{t \in \mathcal{T}} I(U; Z_t^n) \big)$$

where the random variables $U - X^n - (Y_t^n, Z_t^n)$ form a Markov chain.

In addition, it is straight forward to show, similarly as in Theorem 3, that for the compound wiretap channel with side information, the optimal code must have the vanishing output variation property.

Corollary 2: Let $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ be a sequence of wiretap codes achieving the secrecy capacity of the compound wiretap channel with side information. Let and $E : \mathcal{J}_n \longrightarrow \mathcal{P}(\mathcal{X}^n)$ be the corresponding stochastic encoder. Then, there exists for all $t \in \mathcal{T}$ measures ϑ_t on $\mathcal{Z}^n$ and some $\beta > 0$ such that for all $j \in \mathcal{J}_n$ it holds $\|\overline{V}_t^n(\cdot|j) - \vartheta_t\| \leq 2^{-n\beta}$. This means, the optimal code has the almost vanishing output variation property, cf. Definition 7.

B. Multiple Wiretappers

Finally, we want to outline how the results derived for the compound wiretap channel immediately yield solutions for the *multicast channel with multiple wiretappers*. In this scenario, the transmitter wants to transmit a common (multicast) message to group of legitimate receivers while keeping the information secret from a group of nonlegitimate users as depicted in Fig. 1. The concept of compound wiretap channels provides a framework which also includes this scenario. In this case, the number of possible channel realizations corresponds to the number of legitimate users and wiretappers respectively. Thus, each channel realization can be interpreted as a certain legitimate receiver or wiretapper.

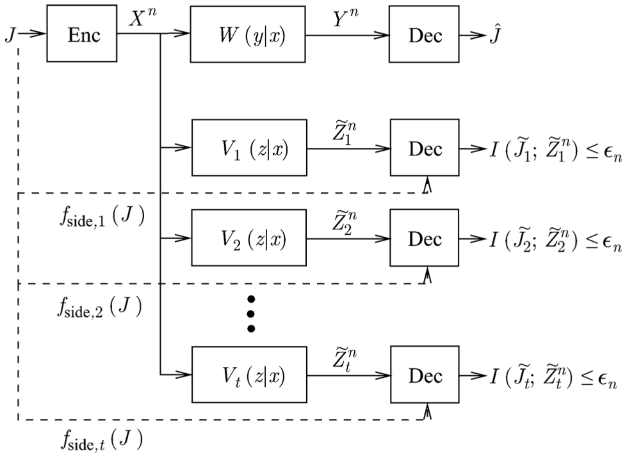


Fig. 4. Wiretap channel with multiple wiretappers. Each wiretapper has his own side information $f_{\text{side},t}(J)$ available, which restricts the message to a certain subset $\widetilde{\mathcal{J}}_t \subseteq \mathcal{J}_n$ with $|\widetilde{\mathcal{J}}_t| \geq 2$.

Let us start with the scenario with only one legitimate user but a group of wiretappers to be kept secret. Then the corresponding compound wiretap channel, cf. Definition 6, becomes

$$\mathfrak{W} = \big\{(W, V_t) : t \in \mathcal{T}\big\}$$

and each channel realization V_t, $t \in \mathcal{T}$, corresponds to another wiretapper which has to be kept secret. Intuitively, it is clear that keeping one wiretapper with an unknown channel realization ignorant is the same task as keeping a whole group of wiretappers with known channel realizations ignorant. The only difference is that in the case of multiple wiretappers, each wiretapper may have different side information about the transmitted message available. This is called the *wiretap channel with multiple wiretappers* and visualized in Fig. 4.

The criteria of strong secrecy and maximum uncertainty for compound wiretap channels already incorporate all possible side information, cf. Equation (19) and (20). In more detail, for strong secrecy we require $I(\widetilde{J}; \widetilde{Z}_t^n)$ to be small for all $t \in \mathcal{T}$ and, more important, for all possible side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n$, $|\widetilde{\mathcal{J}}| \geq 2$. Similarly, for maximum uncertainty we require $\bar{e}_t(\widetilde{\mathcal{J}}) \geq 1 - 1/|\widetilde{\mathcal{J}}| - \lambda_n$ for all possible side information $\widetilde{\mathcal{J}} \subseteq \mathcal{J}_n, |\widetilde{\mathcal{J}}| \geq 2$. Thus, this framework immediately captures the scenario with multiple wiretappers with different side information. We obtain from Theorem 6 the following result.

Corollary 3: An achievable secrecy rate for the wiretap channel with multiple wiretappers with side information (for both strong secrecy and maximum uncertainty) is given by

$$R_S = \max_{U - X - (Y, Z_t)} \big(I(U; Y) - \max_{t \in \mathcal{T}} I(U; Z_t) \big)$$

where the random variables $U - X - (Y, Z_t)$ form a Markov chain.

Next, we extend the scenario by transmitting the confidential message not only to one receiver but to a whole group of legitimate receivers, the so called *multicast channel with multiple wiretappers*. Then, the communication scenario is described by the compound wiretap channel as given in Definition 6, where each possible channel realization to the legitimate receiver corresponds to one legitimate receiver of the group and each possible channel realization to the wiretapper to one additional wiretapper. Then, we immediately obtain the following.

Corollary 4: An achievable secrecy rate for the multicast channel with multiple wiretappers with side information (for both strong secrecy and maximum uncertainty) is given by

$$R_S = \max_{U-X-(Y_t,Z_t)} \big(\min_{t\in\mathcal{T}} I(U;Y_t) - \max_{t\in\mathcal{T}} I(U;Z_t) \big)$$

where the random variables $U - X - (Y_t, Z_t)$ form a Markov chain.

VII. Conclusion

Previous works on wiretap channels focused on wiretappers which solely use their received channel output to infer the confidential information. In this paper we studied more powerful wiretappers which have additional side information available for postprocessing. To do so, we considered two different concepts to measure the secrecy of the transmitted message: the information-theoretic-security-based criterion of strong secrecy and the signal-processing-inspired criterion of maximum uncertainty at the wiretapper. We derived the corresponding secrecy capacities and showed that both criteria yield the same secrecy capacity. Moreover, this secrecy capacity equals the secrecy capacity of the classical wiretap channel without side information. Thus, regardless of the considered secrecy criterion, side information at the wiretapper does not affect the performance of the system in terms of secrecy capacity.

In addition to this, we studied the optimal transceiver design which achieves the secrecy capacity. For this, it is shown that the code property of vanishing output variation at the wiretapper plays a crucial role. This property is established as a necessary and sufficient condition to ensure secrecy under the strong secrecy criterion and the maximum uncertainty criterion. Thus, it characterizes the optimal preprocessing at the transmitter. Finally, it is discussed how these results carry over to the practically relevant cases of channel uncertainty as well as multiple legitimate receivers and multiple wiretappers.

Especially the identification of the vanishing output variation property as the optimal preprocessing at the transmitter yields valuable insights why the secrecy capacity of the wiretap channel does not change whether there is side information at the wiretapper or not. Roughly speaking, the vanishing output variation property guarantees that the output at the nonlegitimate wiretapper is always the same regardless of the actual message that is transmitted and how much side information is available at the wiretapper. The side information does not help the wiretapper to extract further information from the received signal. Thus, the received signal at the wiretapper is always useless regardless of the amount of available side information. Consequently, the secrecy capacity does not change whether there is side information or not.

For future work, it will be interesting to extend these results and the obtained insights to more realistic communication scenarios. This includes the impact finite block lengths to characterize the secrecy rate penalty due to finite block lengths compared to the (asymptotic) secrecy capacity result, cf. for instance [21]. But it includes also more realistic channel models which then requires other metrics such as secrecy outage capacity or ergodic secrecy capacity.

Acknowledgment

The authors gratefully acknowledge valuable discussions with the Federal Office for Information Security (BSI) of Germany which raised the problem of side information at wiretappers.

References

[1] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 4-5, pp. 355–580, 2009.

[2] E. A. Jorswieck, A. Wolf, and S. Gerbracht, "Secrecy on the physical layer in wireless networks," *Trends Telecommun. Technol.*, pp. 413–435, Mar. 2010.

[3] R. Liu and W. E. Trappe*, Securing Wireless Communications at the Physical Layer*. New York, NY, USA: Springer, 2010.

[4] M. Bloch and J. Barros*, Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[5] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.

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

[7] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[8] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: the misome wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2010.

[9] A. Khisti and G. 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.

[10] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[11] R. Bustin, R. Liu, H. V. Poor, and S. Shamai, "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun. 2009, pp. 2602–2606.

[12] S. Loyka and C. D. Charalambous, "On optimal signaling over secure MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 443–447.

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

[14] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.*, vol. 4, no. 4, pp. 371–386, 1960.

[15] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP J. Wireless Commun. Netw.*, pp. 1–13, 2009, Article ID 142374.

[16] E. Ekrem and S. Ulukus, "On Gaussian MIMO compound wiretap channels," in *Proc. Conf. Inf. Sciences and Syst.*, Baltimore, MD, USA, Mar. 2010, pp. 1–6.

[17] I. Bjelakovic, H. Boche, and J. Sommerfeld, "Secrecy results for compound wiretap channels," *Prob. Inf. Transmission*, vol. 49, no. 1, pp. 73–98, Mar. 2013.

[18] I. Csiszár, "Almost independence and secrecy capacity," *Prob. Pered. Inform.*, vol. 32, no. 1, pp. 48–57, 1996.

[19] U. M. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," *EUROCRYPT 2000, Lecture Notes in Computer Sci. Springer-Verlag*, vol. 1807, pp. 351–368, May 2000.

[20] I. Csiszár and J. Körner*, Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[21] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

Holger Boche (M'04–SM'07–F'11) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technische Universität Dresden, Dresden, Germany, in 1990 and 1994, respectively. He graduated in mathematics from the Technische Universität Dresden in 1992. From 1994 to 1997, he did postgraduate studies in mathematics at the Friedrich-Schiller Universität Jena, Jena, Germany. He received the Dr.rer.nat. degree in pure mathematics from the Technische Universität Berlin, Berlin, Germany, in 1998.

In 1997, he joined the Heinrich-Hertz-Institut (HHI) für Nachrichtentechnik Berlin, Berlin, Germany. Starting in 2002, he was a Full Professor for mobile communication networks with the Institute for Communications Systems, Technische Universität Berlin. In 2003, he became Director of the Fraunhofer German-Sino Laboratory for Mobile Communications, Berlin, Germany, and in 2004, he became the Director of the Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany. Since October 2010 he has been with the Institute of Theoretical Information Technology and Full Professor at the Technische Universität München, Munich, Germany. He was a Visiting Professor with the ETH Zurich, Zurich, Switzerland, during the 2004 and 2006 Winter terms, and with KTH Stockholm, Stockholm, Sweden, during the 2005 Summer term.

Prof. Boche is a Member of IEEE Signal Processing Society SPCOM and SPTM Technical Committee. He was elected a Member of the German Academy of Sciences (Leopoldina) in 2008 and of the Berlin Brandenburg Academy of Sciences and Humanities in 2009. He received the Research Award "Technische Kommunikation" from the Alcatel SEL Foundation in October 2003, the "Innovation Award" from the Vodafone Foundation in June 2006, and the Gottfried Wilhelm Leibniz Prize from the Deutsche Forschungsgemeinschaft (German Research Foundation) in 2008. He was corecipient of the 2006 IEEE Signal Processing Society Best Paper Award and recipient of the 2007 IEEE Signal Processing Society Best Paper Award.

Rafael F. Schaefer (formerly Wyrembelski) (S'08–M'12) received the Dipl.-Ing. degree in electrical engineering and computer science, in 2007, from the Technische Universität Berlin, Germany, and the Dr.-Ing. degree in electrical engineering, in 2012, from the Technische Universität München.

Between 2007 and 2010, he worked as a research and teaching assistant at the Heinrich-Hertz-Lehrstuhl für Mobilkommunikation at the Technische Universität Berlin, Germany. Since November 2010, he has been with the Lehrstuhl für Theoretische Informationstechnik at the Technische Universität München, Germany, where he is currently working as a Postdoctoral researcher.