Uncertain Wiretap Channels and Secure Estimation

Moritz Wiese Karl Henrik Johansson Tobias J. Oechtering
Panos Papadimitratos Henrik Sandberg Mikael Skoglund

May 3, 2016

Abstract

Uncertain wiretap channels are introduced. Their zero-error secrecy capacity is defined. If the sensor-estimator channel is perfect, it is also calculated. Further properties are discussed. The problem of estimating a dynamical system with nonstochastic disturbances is studied where the sensor is connected to the estimator and an eavesdropper via an uncertain wiretap channel. The estimator should obtain a uniformly bounded estimation error whereas the eavesdropper's error should tend to infinity. It is proved that the system can be estimated securely if the zero-error capacity of the sensor-estimator channel is strictly larger than the logarithm of the system's unstable pole and the zero-error secrecy capacity of the uncertain wiretap channel is positive.

1 Introduction

If "independent noise" is assumed for every time step, it tends to be considered as stochastic in information theory. In contrast to this, in robust control theory it is common to consider dynamical systems with nonstochastic disturbances. In order to give a unified framework for the latter case, Nair has proposed a "nonstochastic information theory" [1]. The basic channel model in [1] is the newly introduced *uncertain channel*, a rule which determines which channel input can generate which channel outputs *without* weighting the possible outputs given the inputs. For finite alphabets, every uncertain channel thus corresponds to a 0-1-matrix obtained from a stochastic matrix by replacing every positive entry by 1. Thus, uncertain channels are natural objects in zero-error information theory. Nair also introduced an analog to mutual information which plays the same role for the zero-error capacity of uncertain channels as mutual information for the capacity of discrete memoryless channels.

In [1], Nair applied his nonstochastic information theory to the problem of estimating an unstable scalar dynamical system with nonstochastic disturbances at a remote location which obtains sensor data through an uncertain channel $\mathbf{T}$. He showed that the estimation

All authors are with the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden.
E-mails: {moritzw, kallej, oech, papadim, hsan, skoglund}@kth.se

error can be bounded uniformly if the zero-error capacity $C_0(\mathbf{T})$ of $\mathbf{T}$ is strictly larger than the logarithm of the system's unstable pole $\lambda > 1$. This is "almost" necessary as well in the sense that $C_0(\mathbf{T}) \geq \log \lambda$ is required for uniform boundedness of the estimation error.

In this paper we add to the above problem that an eavesdropper overhears the communication between sensor and estimator via a second uncertain channel. The estimation error at the intended location should again be bounded uniformly, whereas for every eavesdropper output sequence there should be two system paths whose distance tends to infinity with increasing time. We call this the problem of secure estimation. A similar problem has been studied in [2] for the case of stochastic system and channel noise.

For our nonstochastic setting, this leads to the introduction of the uncertain wiretap channel: a pair $(\mathbf{T}_B, \mathbf{T}_C)$ of uncertain channels with common input alphabet. A zero-error wiretap code is a zero-error code for $\mathbf{T}_B$ such that every eavesdropper output word can be generated by at least two different messages. Surprisingly, positivity of the zero-error secrecy capacity $C_0(\mathbf{T}_B, \mathbf{T}_C)$ is sufficient, in addition to Nair's sufficient condition for a bounded estimation error, in order for secure estimation to be possible. The reason for this is that the system's instability helps to achieve the goal of security as soon as a sufficiently large error on the eavesdropper side has been introduced at the beginning of transmission.

The schemes for data transmission from the sensor to the estimator apply block codes. Thus there are inter-decoding times where no new data arrive at the estimator. The error at those times increases with communication delay. On the other hand, we show that the estimation error at decoding times can be made to vanish asymptotically at the cost of increased delay. Similarly, we provide a lower bound on the speed of divergence for the eavesdropper's error which increases with increasing delay.

We calculate the secrecy capacity in the case of a perfect sensor-estimator channel. It either equals zero or the logarithm of the size of the input alphabet. An example shows that for general uncertain wiretap channels, no secure message transmission may be possible at blocklength 1, whereas a positive transmission rate is achieved for blocklengths ≥ 2. It also shows that encoders for zero-error wiretap codes in general have to be strictly uncertain channels, i. e. every message can be mapped to several possible codewords similar to the use of stochastic encoders for stochastic wiretap channels. We do not apply Nair's nonstochastic information-theoretic quantities in any of the analyses. Further, uncertain wiretap channels do not appear to provide new insights for the study of zero-error capacity, an overview of which is given in [3].

Outline: Section II describes the problems considered, Section III presents the results and Section IV contains the proofs.

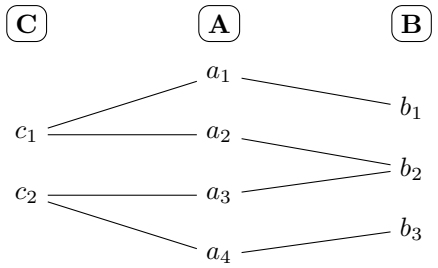

Figure 1: The channel from Example 2. A line between a_i and b_j indicates that $b_j \in \mathbf{T}_B(a_i)$, similar for a_i and c_j.

2 Model

2.1 Uncertain Channels

Let $\mathbf{A}, \mathbf{B}$ be finite alphabets. An *uncertain channel from* $\mathbf{A}$ *to* $\mathbf{B}$ is a mapping $\mathbf{T} : \mathbf{A} \to 2^{\mathbf{B}}_* := 2^{\mathbf{B}} \setminus \{\varnothing\}$. For any $a \in \mathbf{A}$, the set $\mathbf{T}(a)$ is the family of possible output values of the channel given the input a. Only one of the elements of $\mathbf{T}(a)$ will actually be attained when transmitting a. That $\mathbf{T}(a) \neq \varnothing$ for all a means that every input generates an output. We will write $\mathrm{ran}(\mathbf{T})$ for the set of possible outputs of $\mathbf{T}$, i. e. $\mathrm{ran}(\mathbf{T}) = \cup_{a \in \mathbf{A}} \mathbf{T}(a)$.

An *M-code* is a collection $\{\mathbf{F}(m) : 1 \leq m \leq M\}$ of nonempty and mutually disjoint subsets of $\mathbf{A}$. This is equivalent to an uncertain channel $\mathbf{F} : \{1, \ldots, M\} \to 2^{\mathbf{A}}_*$ with disjoint output sets, so we will often denote such a code by $\mathbf{F}$. The necessity of codes with $|\mathbf{F}(m)| \geq 2$ for some m is shown in Examples 1 and 2. It is similar to the necessity of stochastic encoders for stochastic wiretap channels.

Obviously, first applying $\mathbf{F}$ and then $\mathbf{T}$ leads to a new uncertain channel $\mathbf{T} \circ \mathbf{F} : \{1, \ldots, M\} \to 2^{\mathbf{B}}_*$ called the *composition of* $\mathbf{F}$ *and* $\mathbf{T}$. Formally, we have for any $m \in \{1, \ldots, M\}$

$$(\mathbf{T} \circ \mathbf{F})(m) := \mathbf{T}(\mathbf{F}(m)) := \bigcup_{a \in \mathbf{F}(m)} \mathbf{T}(a).$$

A nonstochastic M-code $\mathbf{F}$ is called a *zero-error M-code for* $\mathbf{T}$ if for any $m, m' \in \{1, \ldots, M\}$ with $m \neq m'$

$$\mathbf{T}(\mathbf{F}(m)) \cap \mathbf{T}(\mathbf{F}(m')) = \varnothing. \tag{1}$$

Thus every possible channel output $y \in \mathrm{ran}(\mathbf{T} \circ \mathbf{F})$ can be associated to a unique message m. For this to hold it is necessary that the sets $\mathbf{F}(m)$ be disjoint, which is the reason for this assumption in the definition of M-codes.

Given an additional finite alphabet $\mathbf{C}$, an *uncertain wiretap channel* is a pair of uncertain channels $(\mathbf{T}_B : \mathbf{A} \to 2^{\mathbf{B}}_*, \mathbf{T}_C : \mathbf{A} \to 2^{\mathbf{C}}_*)$. The interpretation is that the outputs of channel $\mathbf{T}_B$ are received by the intended receiver, whereas the outputs of $\mathbf{T}_C$ are heard by an eavesdropper. See Fig. 1 for an example of an uncertain wiretap channel.

An M-code $\mathbf{F}$ is called a *zero-error wiretap M-code for* $(\mathbf{T}_B, \mathbf{T}_C)$ if it is a zero-error code for $\mathbf{T}_B$ and additionally for every $c \in \mathrm{ran}(\mathbf{T}_C \circ \mathbf{F})$ there are messages $m \neq m'$ such that

$$c \in \mathbf{T}_C(\mathbf{F}(m)) \cap \mathbf{T}_C(\mathbf{F}(m')). \tag{2}$$

Thus every output $c \in \mathrm{ran}(\mathbf{T}_C \circ \mathbf{F})$ can be generated by at least two possible messages.

We define the *n-fold product of an uncertain channel* $\mathbf{T} : \mathbf{A} \to 2^{\mathbf{B}}_*$ as the uncertain channel $\mathbf{T}^n : \mathbf{A}^n \to (2^{\mathbf{B}}_*)^n$ defined by

$$\mathbf{T}^n(a^n) = \mathbf{T}(a_1) \times \cdots \times \mathbf{T}(a_n). \tag{3}$$

We call an M-code $\mathbf{F}$ on the alphabet $\mathbf{A}^n$ an (M, n)-*code*. Given an uncertain channel $\mathbf{T}$, an (M, n)-code $\mathbf{F}$ is called a *zero-error (M, n)-code for* $\mathbf{T}$ if (1) is satisfied with $\mathbf{T} \circ \mathbf{F}$ replaced by $\mathbf{T}^n \circ \mathbf{F}$. We set $N_{\mathbf{T}}(n)$ to be the maximal M such that there exists a zero-error (M, n)-code for $\mathbf{T}$ and define the *zero-error capacity of* $\mathbf{T}$ by

$$C_0(\mathbf{T}) := \sup_n \frac{\log N_{\mathbf{T}}(n)}{n}. \tag{4}$$

Given an uncertain wiretap channel $(\mathbf{T}_B, \mathbf{T}_C)$, an (M, n)-code $\mathbf{F}$ is called a *zero-error wiretap (M, n)-code for* $(\mathbf{T}_B, \mathbf{T}_C)$ if it is a zero-error code for $\mathbf{T}_B$ and if (2) holds with $\mathbf{T}_C \circ \mathbf{F}$ replaced by $\mathbf{T}_C^n \circ \mathbf{F}$. We define $N_{(\mathbf{T}_B, \mathbf{T}_C)}(n)$ to be the maximal M such that there exists a zero-error wiretap (M, n)-code for $(\mathbf{T}_B, \mathbf{T}_C)$. Then

$$C_0(\mathbf{T}_B, \mathbf{T}_C) := \sup_n \frac{\log N_{(\mathbf{T}_B, \mathbf{T}_C)}(n)}{n} \tag{5}$$

is called the *zero-error secrecy capacity of* $(\mathbf{T}_B, \mathbf{T}_C)$. Due to the superadditivity of the sequences $\log N_{\mathbf{T}}(n)$ and $\log N_{(\mathbf{T}_B, \mathbf{T}_C)}(n)$, the suprema in (4) and (5) can be replaced by limits by the well-known Fekete's lemma [4], see also [5].

2.2 The Unstable Dynamical System

Let $\lambda > 1$ and consider the real-valued system

$$x(t+1) = \lambda x(t) + w(t), \tag{6a}$$

$$x(0) = 0. \tag{6b}$$

where $w(t)$ is a nonstochastic disturbance with range $[-\Omega/2, \Omega/2]$ for some $\Omega > 0$. With

$$\tilde{\Omega}^{(t)} := \frac{\Omega}{\lambda - 1}(\lambda^t - 1), \tag{7}$$

the range of possible values of this system at time t equals $[-\tilde{\Omega}^{(t)}/2, \tilde{\Omega}^{(t)}/2]$, whose diameter grows exponentially in t. A sensor performs perfect state measurements, encodes them and sends them through an uncertain wiretap channel $(\mathbf{T}_B, \mathbf{T}_C)$. The dynamic system and the

channel are synchronous, i. e. one symbol can be transmitted through the channel at every system time step. The goal is that the receiver of $\mathbf{T}_B$ (the estimator) be able to estimate the state with bounded estimation error and the eavesdropper's estimation error tend to infinity.

Formally, a *transmission scheme* $(n_k, f_k, \varphi_k)_{k=1}^{\infty}$ consists of a bounded sequence of positive natural numbers $(n_k)_{k=1}^{\infty}$ and, defining $t_k := \sum_{i=1}^{k} n_i$, for each $k \in \mathbb{N}$ an uncertain channel $f_k : \mathbb{R}^{t_k} \to 2_*^{\mathbf{X}^{n_k}}$ and a mapping $\varphi_k : \mathbf{Y}^{t_k} \to \mathbb{R}^{n_k}$. Every uncertain channel f_k maps the observations of the system state up till time t_k into one of several possible codewords of length n_k. The receiver of $\mathbf{T}_B$ uses φ_k to produce from all symbols received so far an estimate $\hat{x}(t_k), \ldots, \hat{x}(t_{k+1} - 1)$ of the system states $x(t_k), \ldots, x(t_{k+1} - 1)$.

The minimal delay which has to be tolerated is $\max_k n_k$. At this delay, the receiver has good estimates for the states at times t_k but has to extrapolate for the states $x(t_k + 1), \ldots, x(t_{k+1} - 1)$. In particular, for the first $t_1 - 1$ steps of the evolution, the estimator has to rely on a rule which is independent of any observations and which we assume to estimate $\hat{x}(t) = 0$ $(0 \leq t \leq t_1 - 1)$. Further, every system path $(x(t))_{t=0}^{\infty}$ generates a sequence $(c_t)_{t=1}^{\infty}$ of eavesdropper outputs.

Given a transmission scheme $(n_k, f_k, \varphi_k)_{k=1}^{\infty}$ and a sequence of estimates $(\hat{x}(t))_{t=0}^{\infty}$, we denote by $\mathbf{R}_B((\hat{x}(t))_{t=0}^{\infty})$ the set of system paths $(x(t))_{t=0}^{\infty}$ which using the transmission scheme can generate $(\hat{x}(t))_{t=0}^{\infty}$. One can consider $\mathbf{R}_B$ as an uncertain channel in the reverse direction with $\mathbb{R}^{\infty}$ as input and output alphabet. Similarly, for any infinite sequence $(c_t)_{t=1}^{\infty} \in \mathbf{Z}^{\infty}$ of eavesdropper outputs, we denote by $\mathbf{R}_C((c_t)_{t=1}^{\infty})$ the set of system paths $(x(t))_{t=0}^{\infty}$ which can give rise to $(c_t)_{t=1}^{\infty}$.

For two sequences $(a_t)_{t=1}^{\infty}$ and $(b_t)_{t=1}^{\infty}$ let us define their distance to be $\|(a_t) - (b_t)\|_{\infty} := \sup_t |a_t - b_t|$. For a set S of sequences we define its diameter by

$$\operatorname{diam}(S) := \sup\{\|(a_t) - (b_t)\|_{\infty} : (a_t), (b_t) \in S\}.$$

The transmission scheme $(n_k, f_k, \varphi_k)_{k=1}^{\infty}$ is called *reliable* if the estimation error is bounded uniformly in the estimates, i. e. there exists a constant $\kappa > 0$ such that for every possible estimate sequence $(\hat{x}(t))_{t=0}^{\infty}$,

$$\sup\{\|(x(t)) - (\hat{x}(t))\|_{\infty} : (x(t))_{t=0}^{\infty} \in \mathbf{R}_B((\hat{x}(t))_{t=0}^{\infty})\} \leq \kappa.$$

Further, $(n_k, f_k, \varphi_k)_{k=1}^{\infty}$ is called *secure* if for every sequence $(c_t)_{t=1}^{\infty} \subset \mathbf{C}^{\infty}$

$$\operatorname{diam}(\mathbf{R}_C((c_t)_{t=1}^{\infty})) = \infty.$$

Note that security is an asymptotic property due to the boundedness of the range of possible system states in any finite time horizon, cf. (7). Upon receiving a sequence $(c_t)_{t=1}^{\infty}$ of channel outputs generated by a secure transmission scheme, the eavesdropper will not be able to estimate the system path $(x(t))_{t=0}^{\infty}$ that generated $(c_t)_{t=1}^{\infty}$ with a bounded estimation error.

3 Results

3.1 Main Results

Theorem 1. *A reliable and secure transmission scheme exists if* $C_0(\mathbf{T}_B) > \log\lambda$ *and* $C_0(\mathbf{T}_B, \mathbf{T}_C) > 0$.

The main idea behind Theorem 1 is that the system's instability helps to achieve the goal of security as soon as a sufficiently large error on the eavesdropper side has been introduced at the beginning of transmission.

To apply Theorem 1, $C_0(\mathbf{T}_B)$ and $C_0(\mathbf{T}_B, \mathbf{T}_C)$ have to be known. However, the zero-error capacity $C_0(\mathbf{T}_B)$ is unknown for most channels except a few special cases, cf. [3]. Neither do we provide a general formula for $C_0(\mathbf{T}_B, \mathbf{T}_C)$ here. A solution can be given, though, when the calculation of $C_0(\mathbf{T}_B)$ is trivial.

Theorem 2. *If* $\mathbf{T}_B$ *is an injective function from* $\mathbf{A}$ *to* $\mathbf{B}$*, then* $C_0(\mathbf{T}_B, \mathbf{T}_C) \in \{0, \log|\mathbf{A}|\}$. *Further,* $C_0(\mathbf{T}_B, \mathbf{T}_C) = 0$ *if and only if there is no zero-error wiretap* $(M, 1)$*-code for* $(\mathbf{T}_B, \mathbf{T}_C)$ *for any* $M \geq 2$.

For the proof of Theorem 2, it is sufficient to consider codes with $|\mathbf{F}(m)| = 1$ for all $1 \leq m \leq M$. The number of those elements of $\mathbf{A}^n$ which cannot be used as codewords grows exponentially, at a rate which is less than $\log|\mathbf{A}|$ if and only if there is no zero-error wiretap $(M, 1)$ code for $(\mathbf{T}_B, \mathbf{T}_C)$ for any $M \geq 2$. Thus the number of elements of $\mathbf{A}^n$ that *can* be used either asymptotically grows with rate $\log|\mathbf{A}|$ or equals 0.

3.2 Estimation Error and Divergence Coefficient

We study some additional properties of secure estimation schemes. As mentioned above, using a transmission scheme $(n_k, f_k, \varphi_k)_{k=1}^{\infty}$ with delay $\max_k n_k$, the estimates of system states $x(t)$ with $t \neq t_k$ $(k \in \mathbb{N})$ have to be extrapolated from the last good estimate. Thus the estimation error after a decoding time t_k grows exponentially until the next decoding time t_{k+1}. However, for any $\varepsilon > 0$ the estimation error at times $(t_k)_{k=1}^{\infty}$ can be made smaller than ε at least for large k if the inter-decoding intervals n_k $(k \in \mathbb{N})$ (and thus the inter-decoding estimate errors) are sufficiently large:

Lemma 1. *For every* $\varepsilon > 0$ *there exists a transmission scheme such that for every sequence* $(\hat{x}(t))_{t=0}^{\infty}$ *of estimates and every* $(x(t))_{t=0}^{\infty} \in \mathbf{R}_B((\hat{x}(t))_{t=0}^{\infty})$,

$$\limsup_{k\to\infty} |x(t_k) - \hat{x}(t_k)| \leq \varepsilon.$$

If $C_0(\mathbf{T}_B, \mathbf{T}_C) > \log\lambda$*, then the limit superior can even be replaced by a supremum.*

Another parameter of interest is the speed of divergence of the diameter of the set of possible system states given eavesdropper outputs $(c_t)_{t=1}^{T}$ as $T \to \infty$. Given a zero-error

wiretap (M,n)-code $\mathbf{F}$, we define for every possible eavesdropper channel output $(c_t)_{t=1}^n \in \mathrm{ran}(\mathbf{T}_C^n \circ \mathbf{F})$

$$\delta((c_t)_{t=1}^n) = \max\{|m-m'|+1 : (c_t)_{t=1}^n \in \mathbf{T}_C^n(\mathbf{F}(m)) \cap \mathbf{T}_C^n(\mathbf{F}(m'))\}.$$

Clearly $2 \le \delta((c_t)_{t=1}^n) \le M$. We then set

$$L := \min\{\delta((c_t)_{t=1}^n) : (c_t)_{t=1}^n \in \mathrm{ran}(\mathbf{T}_C^n \circ \mathbf{F})\}$$

and call $\mathbf{F}$ a (M,L,n)-code. We also define

$$\Delta_{(\mathbf{T}_B,\mathbf{T}_C)}(n) := \max\left\{\frac{L-1}{M-1} : \mathbf{F} \text{ is } (M,L,n)\text{-code}\right\}.$$

Clearly, $0 < \Delta_{(\mathbf{T}_B,\mathbf{T}_C)}(n) \le 1$.

Lemma 2. *For every $\varepsilon > 0$ there exists a transmission scheme $(n_k, f_k, \varphi_k)_{k=1}^\infty$ such that for every eavesdropper output sequence $(c_t)_{t=1}^\infty$ there exist system paths $(x(t))_{t=1}^\infty, (x'(t))_{t=1}^\infty \in \mathbf{R}_C((c_t)_{t=1}^\infty)$ satisfying*

$$\liminf_{T\to\infty} \frac{\|(x(t))_{t=1}^T - (x'(t))_{t=1}^T\|_\infty}{\lambda^T} \ge \frac{\Omega}{\lambda-1} \sup_n \Delta_{(\mathbf{T}_B,\mathbf{T}_C)}(n) - \varepsilon.$$

The term on the right-hand side of the inequality in Lemma 2 is positive if ε is chosen small enough. The case $\sup_n \Delta_{(\mathbf{T}_B,\mathbf{T}_C)}(n) = 1$ corresponds to complete eavesdropper ignorance, cf. (7).

3.3 Uncertain Wiretap Channels

We first note that the divergence coefficient increases with increasing blocklength (and hence delay). Thus we find a trade-off between the growth rate for the eavesdropper's estimation error and the delay:

Lemma 3. *If $C_0(\mathbf{T}_B,\mathbf{T}_C) > 0$, then*

$$\sup_n \Delta_{(\mathbf{T}_B,\mathbf{T}_C)}(n) = \lim_{n\to\infty} \Delta_{(\mathbf{T}_B,\mathbf{T}_C)}(n) > 0.$$

Next we have a closer look at the zero-error secrecy capacity of uncertain wiretap channels. To study the zero-error capacity of an uncertain channel $\mathbf{T} : \mathbf{A} \to 2_*^{\mathbf{B}}$, one associates to it the following graph $G(\mathbf{T})$: its vertex set equals $\mathbf{A}$ and an edge is drawn between $a, a' \in \mathbf{A}$ if $\mathbf{T}(a) \cap \mathbf{T}(a') \neq \varnothing$. In that case we write $a \sim a'$.

The graph $G(\mathbf{T}^n)$ corresponding to the n-fold product channel $\mathbf{T}^n$ (see (3)) is the *strong n-fold product of* $G(\mathbf{T})$ denoted by $G(\mathbf{T})^n$, in particular $G(\mathbf{T}^n) = G(\mathbf{T})^n$. Here for any graph G with vertex set $\mathbf{A}$, the strong product G^2 of G with itself is defined as follows: The vertex set of G^2 is $\mathbf{A}^2$ and $(a_1,a_2) \sim (a_1',a_2')$ if 1) $a_1 \sim a_1'$ and $a_2 = a_2'$ or 2) $a_2 \sim a_2'$ and $a_1 = a_1'$ or 3) $a_1 \sim a_1'$ and $a_2 \sim a_2'$.

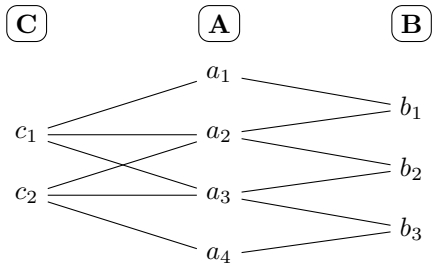

Figure 2: The channel $(\mathbf{T}_B, \mathbf{T}_C)$ from Example 1. A line between a_i and b_j indicates that $b_j \in \mathbf{T}_B(a_i)$, similar for a_i and c_j.

Finding the zero-error capacity of $\mathbf{T}$ now amounts to finding the asymptotic behavior as $n \to \infty$ of the sizes of maximal independent systems of the graphs $G(\mathbf{T}^n)$, cf. [3]. We define an *independent system* in a graph as a set $\{\mathbf{F}(1), \ldots, \mathbf{F}(M)\}$ of mutually disjoint subsets of the vertex set $\mathbf{A}$ such that no two vertices a, a' belonging to different subsets $\mathbf{F}(m) \neq \mathbf{F}(m')$ are connected by an edge.

To treat uncertain wiretap channels $(\mathbf{T}_B, \mathbf{T}_C)$, we consider a hypergraph structure $H(\mathbf{T}_C^n)$ induced on $\mathbf{A}^n$ in addition to the graph structure $G(\mathbf{T}_B^n)$. A hypergraph consists of a vertex set together with a set of subsets, called *hyperedges*, of this vertex set. The vertex set of $H(\mathbf{T}_C^n)$ equals $\mathbf{A}^n$. Every hyperedge is generated by a $(c_t)_{t=1}^n \in \mathbf{C}^n$: we set $e((c_t)_{t=1}^n) := \{(a_t)_{t=1}^n \in \mathbf{A}^n : (c_t)_{t=1}^n \in \mathbf{T}_C^n((a_t)_{t=1}^n)\}$.

It is easy to see that $H(\mathbf{T}_C^n)$ is the *n-fold square product* $H(\mathbf{T}_C)^n$, cf. [6]. For any hypergraph H with vertex set $\mathbf{A}$ and hyperedge set $\mathcal{E} \subset 2^{\mathbf{A}}$, the square product H^2 of H with itself is defined as follows: The vertex set of H^2 is $\mathbf{A}^2$ and the hyperedge set equals $\mathcal{E}^2 := \{e \times e' : e, e' \in \mathcal{E}\}$.

A zero-error wiretap (M, n)-code $\mathbf{F}$ then is nothing but a collection of disjoint subsets $\{\mathbf{F}(1), \ldots, \mathbf{F}(M)\}$ of $\mathbf{A}^n$ satisfying the two following properties:

1. It is an independent system for $G(\mathbf{T}_B^n)$;
2. For every hyperedge e of $H(\mathbf{T}_C^n)$ there exist at least two different m, m' such that e has nonempty intersection with both $\mathbf{F}(m)$ and $\mathbf{F}(m')$.

This (hyper-)graph theoretic language is applied in the proof of Theorem 2. The following very interesting example gives additional insight into the nature of general uncertain wiretap channels and their secrecy capacity.

Example 1. Consider the wiretap channel $(\mathbf{T}_B, \mathbf{T}_C)$ from Fig. 2. $\mathbf{A}$ with $G(\mathbf{T}_B)$ and $H(\mathbf{T}_C)$ is depicted on the left of Fig. 3, $\mathbf{A}^2$ with $G(\mathbf{T}_B^2)$ and $H(\mathbf{T}_C^2)$ on its right. It is easy to check that there is no zero-error wiretap $(M, 1)$-code for any $M \geq 2$. On the other hand, a zero-error wiretap $(4, 2)$-code exists by chooosing the codeword sets as indicated in Fig. 3. Therefore in the general case, in contrast to the situation in Lemma 2, there is no easy

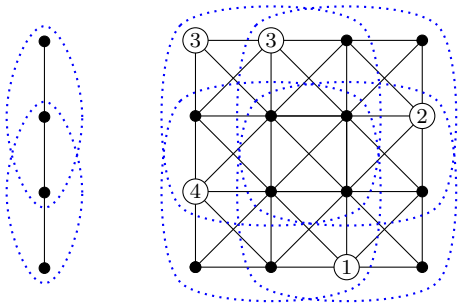

Figure 3: Left: $\mathbf{A}$ with $G(\mathbf{T}_B)$ and $H(\mathbf{T}_C)$. Right: $\mathbf{A}^2$ with $G(\mathbf{T}_B^2)$ and $H(\mathbf{T}_C^2)$. Vertices connected by a solid black line are connected in $G(\mathbf{T}_B)$ or $G(\mathbf{T}_B^2)$, respectively. Vertices within the boundary of a blue dotted line belong to the same hyperedge of $H(\mathbf{T}_C)$ or $H(\mathbf{T}_C^2)$, respectively.

criterion an uncertain wiretap channel satisfies at blocklength 1 if and only if its zero-error secrecy capacity is positive.

This behavior of zero-error wiretap codes for general uncertain wiretap channels is remarkable when it is compared to the behavior of zero-error codes for uncertain channels: An uncertain channel $\mathbf{T}$ has $C_0(\mathbf{T}) > 0$ if and only if there exists an independent system for $G(\mathbf{T})$ with size ≥ 2. Similarly, a stochastic DMC has positive capacity if and only if its blocklength-1 transmission matrix does not have identical rows. For the secrecy capacity of stochastic wiretap channels, there is van Dijk's criterion [7] for positivity which concerns the blocklength-1 wiretap channel matrix and requires to check a certain function for concavity.

Observe also that in order to obtain a $(4, 2)$-code for the above channel, one message m has to be encoded into a set with $|\mathbf{F}(m)| \geq 2$.

A simpler example illustrating the necessity of codes whose encoding sets $\mathbf{F}(m)$ are not all one-element sets is the following.

Example 2. Consider the wiretap channel shown in Fig. 1. If one only considered codes satisfying $|\mathbf{F}(m)| = 1$ for all messages m, then the maximal M for which a zero-error wiretap M-code exists would be $M = 2$, for example $\mathbf{F} = \{\{a_1\}, \{a_4\}\}$. $M = 4$ is not possible because $\mathbf{T}_B$ can only transmit 3 messages without error. For $M = 3$, either c_1 or c_2 would be generated by only one message.

On the other hand, if one takes the zero-error wiretap code $\mathbf{F} = \{\{a_1\}, \{a_2, a_3\}, \{a_4\}\}$, then three messages can be distinguished at the intended receiver's output and every eavesdropper output is reached by two different messages.

Moreover, examples can be constructed which show the following: If there exists a zero-error wiretap (M, n)-code, then it may be necessary to have codes with $|\mathbf{F}(m)| \geq 2$ to also find a zero-error wiretap (M', n)-code for every $2 \leq M' \leq M$.

4 Proofs

This section contains all the proofs. The first two subsections are devoted to the proof of Theorem 1. Subsection 4.1 contains the quantizer rule applied by the sensor and some basic lemmas which are needed in the analysis of the transmission scheme to be defined. The transmission scheme is defined and analyzed in Subsection 4.2. The proofs of Lemmas 1 and 2 which are based on the transmission scheme defined in Subsection 4.2 are done in Subsection 4.3. The proof of Theorem 2 is contained in Subsection 4.4, followed by the proof of Lemma 3 in Subsection 4.5.

4.1 Proof of Theorem 1: Preliminaries

The first choice to make is the quantizer used by the sensor. For sufficient generality, we assume the rule (6a), but that $x(0) \in I_0$ for some real interval I_0. Let $M \geq 2 \in \mathbb{N}$. Recursively define for $t \geq 1$ and $1 \leq m \leq M$

$$[A(t), B(t)] = \lambda I_{t-1} + \left[-\frac{\Omega}{2}, \frac{\Omega}{2}\right], \tag{8}$$

$$P_{m,t} = A(t) + (B(t) - A(t)) \left[\frac{m-1}{M}, \frac{m}{M}\right], \tag{9}$$

$$m_t = m \qquad \text{if } x(t) \in P_{m,t}, \tag{10}$$

$$I_t = P_{m_t,t}. \tag{11}$$

In the definition of m_t, an uncertain mapping is applied to associate $x(t)$ to one of the two possible values if it lies on the boundary between two partition intervals $P_{m,t}, P_{m+1,t}$.

For every $t \in \mathbb{N}$, the interval I_t is the set of system states which are possible at time t according to the sequence $(m_i)_{i=t}^n$. The interval $[A(t+1), B(t+1)]$ is the set of states the system could be in at time $t+1$ given that its state at time t is contained in I_t. The sets $P_{m,t+1} : 1 \leq m \leq M$ form an equal-sized partition of $[A(t+1), B(t+1)]$, and m_{t+1} is the index of the partition atom actually containing the system state. Clearly, every path $(x(t))_{t=0}^\infty$ generates an infinite sequence $(m_t)_{t=1}^\infty$.

The next lemma is needed in the analysis of the intended receiver's estimation error and proved by induction over the recursion (8)-(11).

Lemma 4. *For every $t \in \mathbb{N}$ and $1 \leq m \leq M$,*

$$|P_{m,t}| = \begin{cases} \left(\frac{\lambda}{M}\right)^t \left(|I_0| - \frac{\Omega}{M-\lambda}\right) + \frac{\Omega}{M-\lambda} & \text{if } \lambda \neq M, \\ |I_0| + t\frac{\Omega}{M} & \text{if } \lambda = M. \end{cases} \tag{12}$$

In particular, $\sup_t |I_t| < \infty$ if and only if $\lambda < M$. In that case

$$\sup_t |I_t| = \max\left\{|I_0|, \frac{\Omega}{M-\lambda}\right\}.$$

Proof. Write $I_t := [I_{t,\min}, I_{t,\max}]$ for every $t \in \mathbb{N}$. Then note that by (8)

$$[A(t+1), B(t+1)] = \left[\lambda I_{t,\min} - \frac{\Omega}{2}, \lambda I_{t,\max} + \frac{\Omega}{2}\right], \tag{13}$$

which implies that $B(t+1) - A(t+1) = \lambda|I_t| + \Omega$. Hence by (11) and (9)

$$|P_{m,t+1}| = \frac{B(t+1) - A(t+1)}{M} = \frac{\lambda}{M}|I_t| + \frac{\Omega}{M}. \tag{14}$$

Next (12) is established by induction over t, using (14). It is simple if $\lambda = M$. Now assume $\lambda \neq M$. Clearly the statement if true for $t = 0$. Using the induction hypothesis and (14), we then get

$$\begin{aligned}
|P_{m,t+1}| &= \frac{\lambda}{M}|I_t| + \frac{\Omega}{M} \\
&= \frac{\lambda}{M}\left(\left(\frac{\lambda}{M}\right)^t \left(|I_0| - \frac{\Omega}{M-\lambda}\right) + \frac{\Omega}{M-\lambda}\right) + \frac{\Omega}{M} \\
&= \left(\frac{\lambda}{M}\right)^{t+1} \left(|I_0| - \frac{\Omega}{M-\lambda}\right) + \frac{\Omega}{M-\lambda}\left(\frac{\lambda}{M} + \frac{M-\lambda}{M}\right) \\
&= \left(\frac{\lambda}{M}\right)^{t+1} \left(|I_0| - \frac{\Omega}{M-\lambda}\right) + \frac{\Omega}{M-\lambda}.
\end{aligned}$$

This completes the proof. □

Denote by $\hat{x}(t)$ the mid point of I_t for $t \geq 0$. For the analysis of the diameter of the set of paths compatible with the eavesdropper's outputs, we first derive a recursion formula for the sequence $(\hat{x}(t))_{t=1}^{\infty}$ given a sequence of partition indices $(m_t)_{t=1}^{\infty}$.

Lemma 5. *Let $M \in \mathbb{N}$ and for every t let $1 \leq m_t \leq M$. Let $\hat{x}(t)$ be the mid point of $I_t = P_{m_t,t}$ for $t \in \mathbb{N}$. Define*

$$\sigma_t := \sum_{i=0}^{t} \left(\frac{\lambda}{M}\right)^t = \frac{M}{M-\lambda}\left(1 - \left(\frac{\lambda}{M}\right)^{t+1}\right).$$

Then for every $t = 0, 1, 2, \ldots$

$$\hat{x}(t) = \lambda^t \left\{\hat{x}(0) - \frac{1}{2}\sum_{i=0}^{t-1}\left(\frac{\Omega\sigma_i}{\lambda^{i+1}} + \frac{|I_0|}{M^i}\right)\left(1 - \frac{2m_{i+1} - 1}{M}\right)\right\}. \tag{15}$$

Proof. For $t \in \mathbb{N}$ and $1 \leq m \leq M$ we set $P_{m,t} = [P_{m,t,\min}, P_{m,t,\max}]$. By definition,

$\hat{x}(t+1) = P_{m_{t+1},t+1,\min} + |P_{1,t+1}|/2$ (using $|P_{1,t+1}| = |P_{m,t+1}|$ for all $1 \le m \le M$). Then

$$\begin{aligned}
\hat{x}(t+1) &= A(t+1) + (m_{t+1} - \frac{1}{2})|P_{1,t+1}| && \text{by (9)} \\
&= \lambda P_{m_t,t,\min} - \frac{\Omega}{2} + (m_{t+1} - \frac{1}{2})|P_{1,t+1}| && \text{by (8)} \\
&= \lambda \hat{x}(t) - \frac{\lambda|P_{1,t}|}{2} - \frac{\Omega}{2} + (m_{t+1} - \frac{1}{2})|P_{1,t+1}| && \text{by def. of } \hat{x}(t) \\
&= \lambda \hat{x}(t) - \frac{\lambda|P_{1,t}|}{2} - \frac{\Omega}{2} + (m_{t+1} - \frac{1}{2})\left(\frac{\lambda}{M}|P_{1,t}| + \frac{\Omega}{M}\right) && \text{by (14)} \\
&= \lambda \hat{x}(t) - \frac{\lambda|P_{1,t}| + \Omega}{2}\left(1 - \frac{2m_{t+1}-1}{M}\right). && (16)
\end{aligned}$$

By Lemma 4 and (16)

$$\begin{aligned}
\hat{x}(t+1) &= \lambda \hat{x}(t) - \left(\frac{\lambda^{t+1}|I_0|}{2M^t} - \frac{\lambda^{t+1}}{2M^t}\frac{\Omega}{M-\lambda} + \frac{\lambda}{2}\frac{\Omega}{M-\lambda} + \frac{\Omega}{2}\right)\left(1 - \frac{2m_{t+1}-1}{M}\right) \\
&= \lambda \hat{x}(t) - \left(\frac{\lambda^{t+1}|I_0|}{2M^t} - \frac{\Omega}{2}\frac{\lambda^{t+1} - \lambda M^t - (M-\lambda)M^t}{M^t(M-\lambda)}\right)\left(1 - \frac{2m_{t+1}-1}{M}\right) \\
&= \lambda \hat{x}(t) - \left(\frac{\lambda^{t+1}|I_0|}{2M^t} - \frac{\Omega}{2}\frac{M}{M-\lambda}\frac{\lambda^{t+1} - M^{t+1}}{M^{t+1}}\right)\left(1 - \frac{2m_{t+1}-1}{M}\right) \\
&= \lambda \hat{x}(t) - \left(\frac{\lambda^{t+1}}{M^t}\frac{|I_0|}{2} + \frac{\Omega}{2}\sigma_t\right)\left(1 - \frac{2m_{t+1}-1}{M}\right). && (17)
\end{aligned}$$

Now we use induction to prove the claim. It is certainly correct for $t = 0$. Assume the claim has been proven for all integers up to t. We obtain from (17)

$$\begin{aligned}
\hat{x}(t+1) &= \lambda^{t+1}\left\{\hat{x}(0) - \frac{1}{2}\sum_{i=0}^{t-1}\left(\frac{\Omega\sigma_i}{\lambda^{i+1}} + \frac{|I_0|}{M^i}\right)\left(1 - \frac{2m_{i+1}-1}{M}\right)\right\} \\
&\qquad - \frac{\lambda^{t+1}}{2}\left\{\left(\frac{\Omega\sigma_t}{\lambda^{t+1}} + \frac{|I_0|}{M^t}\right)\left(1 - \frac{2m_{t+1}-1}{M}\right)\right\} \\
&= \lambda^{t+1}\left\{\hat{x}(0) - \frac{1}{2}\sum_{i=0}^{t}\left(\frac{\Omega\sigma_i}{\lambda^{i+1}} + \frac{|I_0|}{M^i}\right)\left(1 - \frac{2m_{i+1}-1}{M}\right)\right\}.
\end{aligned}$$

This completes the proof. □

Next assume that we have two systems obeying (6a). The paths of one of them start in an interval I_0 and those of the other in an interval I_0' with $|I_0| = |I_0'|$. The same quantizer rules (8)-(11) are applied for both systems, generating sequences $(m_t)_{t=1}^{\infty}, (I_t)_{t=0}^{\infty}$ and $(m_t')_{t=1}^{\infty}, (I_t')_{t=0}^{\infty}$, respectively. For $t \ge 0$, denote by $\hat{x}(t)$ the mid point of I_t and by $\hat{x}'(t)$ that of I_t'. The next two lemmas will be used in the security analysis of the scheme we are going to define.

Lemma 6. *Let $L, M \ge 2$ and for every t let $1 \le m_t' < m_t \le M$ with $m_t - m_t' \ge L - 1$. Then*

$$\liminf_{t\to\infty} \frac{\hat{x}(t) - \hat{x}'(t)}{\lambda^t} \ge \hat{x}(0) - \hat{x}'(0) + \frac{L-1}{M-1}\left(\frac{\Omega}{\lambda - 1} + |I_0|\right).$$

Proof. By Lemma 5, for any $n \in \mathbb{N}$,

$$\hat{x}(t) - \hat{x}'(t) = \lambda^t \Bigg(\hat{x}(0) - \hat{x}'(0) + \frac{\Omega}{M} \sum_{i=0}^{t-1} \frac{\sigma_i}{\lambda^{i+1}} (m_{i+1} - m'_{i+1}) + |I_0| \sum_{i=0}^{t-1} \frac{1}{M^{i+1}} (m_{i+1} - m'_{i+1}) \Bigg). \tag{18}$$

Observe that

$$\frac{\sigma_i}{\lambda^{i+1}} = \frac{M}{M-\lambda} \left(\frac{1}{\lambda^{i+1}} - \frac{1}{M^{i+1}} \right) \tag{19}$$

and recall that $m_t - m'_t \geq L - 1$ for every t. Hence (18) can be lower-bounded by

$$\begin{aligned} &\lambda^t \Bigg(\hat{x}(0) - \hat{x}'(0) + \frac{\Omega(L-1)}{M-\lambda} \sum_{i=0}^{t-1} \left(\frac{1}{\lambda^{i+1}} - \frac{1}{M^{i+1}} \right) + |I_0|(L-1) \sum_{i=0}^{t-1} \frac{1}{M^{i+1}} \Bigg) \\ &= \lambda^t \Bigg(\hat{x}(0) - \hat{x}'(0) + \frac{\Omega(L-1)}{M-\lambda} \left(\frac{1-\lambda^{-t}}{\lambda - 1} - \frac{1-M^{-t}}{M-1} \right) + |I_0|(L-1) \frac{1-M^{-t}}{M-1} \Bigg). \end{aligned}$$

The theorem is proven once one observes that as $t \to \infty$,

$$\begin{aligned} &\frac{\Omega(L-1)}{M-\lambda} \left(\frac{1-\lambda^{-t}}{\lambda - 1} - \frac{1-M^{-t}}{M-1} \right) + |I_0|(L-1) \frac{1-M^{-t}}{M-1} \\ &\longrightarrow \frac{\Omega(L-1)}{M-\lambda} \left(\frac{1}{\lambda - 1} - \frac{1}{M-1} \right) + |I_0|(L-1) \frac{1}{M-1} \\ &= \frac{\Omega(L-1)}{M-\lambda} \frac{M-\lambda}{(M-1)(\lambda-1)} + |I_0|(L-1) \frac{1}{M-1} \\ &= \frac{L-1}{M-1} \left(\frac{\Omega}{\lambda - 1} + |I_0| \right). \end{aligned}$$

□

Lemma 7. *Let $M \in \mathbb{N}$ and for every t let $1 \leq m_t, m'_t \leq M$. If*

$$|\hat{x}(0) - \hat{x}'(0)| > \frac{\Omega}{\lambda - 1} + |I_0|, \tag{20}$$

then for every $t = 1, 2, \ldots$

$$\liminf_{t \to \infty} \frac{|\hat{x}(t) - \hat{x}'(t)|}{\lambda^t} \geq |\hat{x}(0) - \hat{x}'(0)| - \frac{\Omega}{\lambda - 1} - |I_0|.$$

Proof. By Lemma 4,

$$|\hat{x}(t) - \hat{x}'(t)| = \lambda^t \left| (\hat{x}(0) - \hat{x}'(0)) + \frac{1}{M} \sum_{i=0}^{t-1} \left(\frac{\Omega \sigma_i}{\lambda^{i+1}} + \frac{|I_0|}{M^i} \right) (m_{i+1} - m'_{i+1}) \right|. \tag{21}$$

By the triangle inequality, the absolute value term on the right-hand side of (21) is lower-bounded by

$$\left| |\hat{x}(0) - \hat{x}'(0)| - \left| \frac{1}{M} \sum_{i=0}^{t-1} \left(\frac{\Omega \sigma_i}{\lambda^{i+1}} + \frac{|I_0|}{M^i} \right) (m_{i+1} - m'_{i+1}) \right| \right|. \tag{22}$$

Using (19),

$$\left|\frac{1}{M}\sum_{i=0}^{t-1}\left(\frac{\Omega\sigma_i}{\lambda^{i+1}}+\frac{|I_0|}{M^i}\right)(m_i-m_i')\right| \tag{23}$$

$$\leq \frac{M-1}{M}\sum_{i=0}^{t-1}\left(\frac{\Omega M}{M-\lambda}\left(\frac{1}{\lambda^{i+1}}-\frac{1}{M^{i+1}}\right)+\frac{|I_0|}{M^i}\right)$$

$$=(M-1)\left\{\frac{\Omega}{M-\lambda}\frac{1-\lambda^{-t}}{\lambda-1}+\left(|I_0|-\frac{\Omega}{M-\lambda}\right)\frac{1-M^{-t}}{M-1}\right\}$$

$$=\frac{\Omega}{M-\lambda}\frac{M-1}{\lambda-1}\left(1-\frac{1}{\lambda^t}\right)+\left(|I_0|-\frac{\Omega}{M-\lambda}\right)\left(1-\frac{1}{M^t}\right). \tag{24}$$

As t tends to infinity, (24) converges to

$$\frac{\Omega}{M-\lambda}\left(\frac{M-1}{\lambda-1}-1\right)+|I_0|=\frac{\Omega}{\lambda-1}+|I_0|.$$

This proves the lemma. □

4.2 Proof of Theorem 1: Transmission Schemes

For any $n \geq 1$, let us introduce the *n-sampled system*

$$x^{(n)}(k+1)=\lambda^n x^{(n)}(k)+w^{(n)}(k),$$
$$x^{(n)}(0)=0,$$

where $w^{(n)}(k)$ is a nonstochastic disturbance in the range $[-\tilde{\Omega}^{(n)}/2, \tilde{\Omega}^{(n)}/2]$ (cf. (7)). The n-sampled system describes the system (6) at the points $0, n, 2n, \ldots$

Let us first assume that $C_0(\mathbf{T}_B, \mathbf{T}_C) \leq \log\lambda$. Choose n_1, M_1 such that $2 \leq M_1 < \lambda^{n_1}$ and $M_1 \leq N_{(\mathbf{T}_B,\mathbf{T}_C)}(n_1)$ and choose n_2 such that $M_2 := N_{\mathbf{T}_B}(n_2) > \lambda^{n_2}$. Let $L \geq 2$ be chosen such that there exists a zero-error wiretap (M_1, L, n_1)-code $\mathbf{F}$ and let $\mathbf{G}$ be a zero-error (M_2, n_2)-code.

We define a transmission scheme as follows: We do the construction (8)-(11) for the n_1-sampled system with M replaced by M_1 and with $I_0 = \{0\}$, thus obtaining $A^{(n_1)}(k), B^{(n_1)}(k)$, $P_{m,k}^{(n_1)}, m_k, I_k^{(n_1)}$ (omitting the superscript (n_1) at m_k). For some $K \in \mathbb{N}$ to be chosen later and $1 \leq k \leq K$, we set

$$f_k(x(0), \ldots, x(kn_1)) = \mathbf{F}(m_k)$$

The intended receiver uses the mid point $\hat{x}(kn_1)$ of $P_{m_k,k}^{(n_1)}$ as estimate of $x(kn_1)$. For $k > K$, we first define $A^{(n_2)}(k-K), B^{(n_2)}(k-K), P_{m,k-K}^{(n_2)}, m_{k-K}, I_{k-K}^{(n_2)}$ as in (8)-(11) but with $I_0 = P_{m_K,K}^{(n_1)}$ (and again omitting the superscript (n_2) at m_{k-K}). We then set

$$f_k(x(0), \ldots, x(Kn_1+(k-K)n_2) = \mathbf{G}(m_{k-K})$$

Decoding/estimating goes as in the first K steps.

As $M_2 > \lambda^{n_2}$ it is clear that the estimation error for the intended receiver at decoding times $(t_k)_{k=1}^{\infty}$ equals

$$\max\left\{\max_{1\le k\le K}|P^{(n_1)}_{m_k,k}|, \sup_k |P^{(n_2)}_{m_k,k-K}|\right\} < \infty. \tag{25}$$

More precisely, by Lemma 4, the maximum in the curly brackets in (25) equals

$$|P_{m_K,K}| = \left(\left(\frac{\lambda^{n_1}}{M_1}-1\right)^K\right)\frac{\Omega}{\lambda-1}\frac{\lambda^{n_1}-1}{\lambda^{n_1}-M_1} \tag{26}$$

and the supremum inside the curly brackets in (25) equals the maximum of (26) and

$$\frac{\Omega}{\lambda-1}\frac{\lambda^{n_2}-1}{M_2-\lambda^{n_2}}. \tag{27}$$

Thus the intended receiver's estimation error is bounded at decoding times. In between, it can only grow finitely, so the total estimation error is bounded.

To prove security of the transmission scheme defined above, fix an $\varepsilon > 0$. Now assume the eavesdropper receives a channel output sequence $(c_t)_{t=1}^{\infty}$. Lemma 6 implies the existence of paths $(x(t))_{t=0}^{\infty}, (x'(t))_{t=0}^{\infty}$ such that for sufficiently large K, the estimates at time Kn_1 have distance

$$\hat{x}(Kn_1) - \hat{x}'(Kn_1) \ge \lambda^{Kn_1}\left(\frac{L-1}{M_1-1}\frac{\Omega}{\lambda-1} - \varepsilon\right) \tag{28}$$

(note that here, Lemma 6 has to be applied with $|I_0| = 0$ and $\hat{x}(0) = \hat{x}'(0)$). By choosing K even larger if necessary, (20) is satisfied with its left-hand side replaced by $|\hat{x}(Kn) - \hat{x}'(Kn)|$ and the right-hand side by

$$\frac{\tilde{\Omega}^{(n_2)}}{\lambda^{n_2}-1} + |P_{m_K,K}|. \tag{29}$$

This can be seen by applying (28) and by using (26) to show that (29) equals

$$\frac{\tilde{\Omega}^{(n_2)}}{\lambda^{n_2}-1} + |P^{(n_1)}_{m,K}| = \frac{\Omega}{\lambda-1}\left(1 + \left(\left(\frac{\lambda^{n_1}}{M_1}\right)^K - 1\right)\frac{\lambda^{n_1}-1}{M_1-\lambda^{n_1}}\right).$$

One can thus apply Lemma 7 to find that for sufficiently large k (and after enlarging K again if necessary), the distance between $\hat{x}(kn_2)$ and $\hat{x}'(kn_2)$ is lower-bounded by

$$\lambda^{Kn_1+(k-K)n_2}\frac{\Omega}{\lambda-1}\left(\frac{L-1}{M_1-1} - \frac{1}{\lambda^{Kn_1}} - \left(\frac{1}{M_1^K} - \frac{1}{\lambda^{Kn_1}}\right)\frac{\lambda^{n_1}-1}{M_1-\lambda^{n_1}} - 2\varepsilon\frac{\lambda-1}{\Omega}\right). \tag{30}$$

This tends to infinity as $k \to \infty$ and thus proves that the transmission scheme defined satisfies security. We have thus proved that there exists a reliable and secure transmission scheme in the case $C_0(\mathbf{T}_B) > \log\lambda$ and $0 < C_0(\mathbf{T}_B, \mathbf{T}_C) \le \log\lambda$.

Next we treat the case $C_0(\mathbf{T}_B, \mathbf{T}_C) > \log\lambda$. The construction is simpler than the previous case, as it applies the same zero-error wiretap code in every time step. Choose n such that $M := N_{(\mathbf{T}_B,\mathbf{T}_C)}(n) > \lambda^n$. Let $L \ge 2$ be chosen such that there exists a zero-error wiretap (M, L, n)-code $\mathbf{F}$.

We define a transmission scheme as follows: The construction (8)-(11) is done for the n-sampled system with $I_0 = \{0\}$ and thus obtain $A^{(n)}(k), B^{(n)}(k), P^{(n)}_{m,k}, m_k, I^{(n)}_k$ (omitting the superscript (n) at m_k). We then set

$$f_k(x(0), \dots, x(kn)) = \mathbf{F}(m_k).$$

Again, the intended receiver uses the mid point of $P^{(n)}_{m_k,k}$ as estimate of $x(kn)$.

By Lemma 4, the estimation error at times $0, n, 2n, \dots$ is bounded by

$$\frac{\Omega}{\lambda - 1} \frac{\lambda^n - 1}{M - \lambda^n}. \tag{31}$$

Between these times, the error grows, but stays bounded. Hence the total estimation error is bounded, so the above transmission scheme is reliable.

For security, we apply Lemma 6 and find that for any $\varepsilon > 0$, any eavsdropper sequence $(c_t)_{t=1}^{\infty}$) and sufficiently large k, there exist paths $(x(t))_{t=0}^{\infty}, (x'(t))_{t=0}^{\infty} \in \mathbf{R}_C((c_t)_{t=1}^{\infty})$ such that

$$\hat{x}(kn) - \hat{x}'(kn) \geq \lambda^{kn} \left(\frac{\Omega}{\lambda - 1} \frac{L - 1}{M - 1} - \varepsilon \right). \tag{32}$$

Thus the transmission scheme also is secure. Altogether, this proves Theorem 1.

4.3 Proofs of Lemmas 1 and 2

We distinguish the cases $C_0(\mathbf{T}_B, \mathbf{T}_C) \leq \log \lambda$ and $C_0(\mathbf{T}_B, \mathbf{T}_C) > \log \lambda$ and treat both lemmas for each case at once.

Let us start with the case $C_0(\mathbf{T}_B, \mathbf{T}_C) \leq \log \lambda$. The maximal estimation error at decoding times $0, n_1, \dots, Kn_1, Kn_1 + n_2, Kn_1 + 2n_2, \dots$ is given by (25), i. e. the maximum of (26) and (27). The error (26) is obtained at time Kn_1, whereas (27) is the asymptotic error as $k \to \infty$. By choosing n_2 sufficiently large, this asymptotic error can be made arbitrarily small by choice of M_2. Thus for any $\varepsilon > 0$ and for sufficiently large $n_2 = n_2(\varepsilon)$, we obtain

$$|x(Kn_1 + (k - K)n_2) - \hat{x}(Kn_1 + (k - K)n_2)| \leq \varepsilon.$$

This proves Lemma 1 for the case $C_0(\mathbf{T}_B, \mathbf{T}_C) \leq \log \lambda$.

To also show Lemma 2, we just need to have a look at (30). First we choose n_1 so large that

$$\frac{L - 1}{M_1 - 1} \geq \sup_n \Delta_{(\mathbf{T}_B, \mathbf{T}_C)}(n) - \varepsilon.$$

Thus the term in the outer brackets in (30) is lower bounded by

$$\sup_n \Delta_{(\mathbf{T}_B, \mathbf{T}_C)}(n) - \frac{1}{\lambda^{Kn_1}} - \left(\frac{1}{M_1^{Kn_1}} - \frac{1}{\lambda^{Kn_1}} \right) \frac{\lambda^{n_1} - 1}{M_1 - \lambda^{n_1}} - \varepsilon \left(1 + 2 \frac{\lambda - 1}{\Omega} \right)$$

Next with sufficiently large K, it can be ensured that

$$\frac{1}{\lambda^{Kn_1}} + \left(\frac{1}{M_1^{Kn_1}} - \frac{1}{\lambda^{Kn_1}} \right) \frac{\lambda^{n_1} - 1}{M_1 - \lambda^{n_1}} + \varepsilon \left(1 + 2 \frac{\lambda - 1}{\Omega} \right) \leq 2\varepsilon \left(1 + \frac{\lambda - 1}{\Omega} \right). \tag{33}$$

Recall that ε depends on K and can be made arbitrarily small by enlarging K. Hence the term on the right-hand side of (33) can be made arbitrarily small. This proves Lemma 2 for the case $C_0(\mathbf{T}_B, \mathbf{T}_C) \leq \log \lambda$.

We next prove Lemmas 1 and 2 to also hold for the case $C_0(\mathbf{T}_B, \mathbf{T}_C) > \log \lambda$. By (31) and the choice of M, the estimation error at decoding times can be made arbitrarily small by choosing n sufficiently large. Note that this gives the claimed upper bound on the *supremum* of all estimation errors at decoding times. This proves Lemma 1. The proof of Lemma 2 is simple as well because of (32).

4.4 Proof of Theorem 2

For the proof of Theorem 2, observe that one can restrict attention to codes with $|\mathbf{F}(m)| = 1$ because no vertices are connected in $G(\mathbf{T}_B^n)$ for any n. At blocklength n, the only question will be how many elements of $\mathbf{A}^n$ can be used as codewords. We write $a_1^n := (a_1, \ldots, a_n)$ for elements of $\mathbf{A}^n$ and use analogous notation for elements $c_1^n \in \mathbf{C}^n$. It has to be ensured that the eavesdropper cannot infer the codeword a_1^n, and thus the message, from its received $c_1^n \in \mathbf{C}^n$.

To formalize this, we introduce the notion of "subhypergraph" of a hypergraph. Given a hypergraph H with vertex set $\mathbf{V}$ and hyperedge set $\mathcal{E}_H$, we call $\tilde{H}$ a *subhypergraph of* H if the vertex set $\tilde{\mathbf{V}}$ of $\tilde{H}$ is a subset of $\mathbf{V}$ and if each of the hyperedges of $\tilde{H}$ has the form $\tilde{e} = e \cap \tilde{\mathbf{V}}$ for some $e \in \mathcal{E}_H$ (the empty set is not allowed as hyperedge). Obviously, the subhypergraph $\tilde{H}$ is uniquely determined by $\tilde{\mathbf{V}}$ and we denote it by $H|_{\tilde{\mathbf{V}}}$.

Denote by $\mathbf{T}_C^n|_{\mathbf{V}}$ the channel $\mathbf{T}_C^n$ restricted to inputs from $\mathbf{V} \subset \mathbf{A}^n$ and observe that the hypergraph $H(\mathbf{T}_C^n|_{\mathbf{V}})$ is given by the subhypergraph $H(\mathbf{T}_C^n)|_{\mathbf{V}}$ of $H(\mathbf{T}_C^n)$. We can thus formulate our problem by saying that we have to find a large subhypergraph $H^{(n)}$ of $H(\mathbf{T}_C^n)$ which does not contain any hyperedge of cardinality 1.

This subhypergraph is found in several consecutive steps. We set $H(\mathbf{T}_C^n) =: H^{(n)}(0)$. First we eliminate from the possible channel input alphabet $\mathbf{A}^n$ all elements a_1^n which can be uniquely determined by the eavesdropper, i. e. all a_1^n such that $\{a_1^n\}$ is a hyperedge of $H^{(n)}(0)$. If we write

$$\mathbf{A}_1^{(n)}(1) := \{a_1^n \in \mathbf{A}^n : \{a_1^n\} \text{ is a hyperedge of } H^{(n)}(0)\},$$

and $\mathbf{A}_2^{(n)}(1) := \mathbf{A}^n \setminus \mathbf{A}_1^{(n)}(1)$, we thus obtain the subhypergraph $H^{(n)}(1) := H^{(n)}(0)|_{\mathbf{A}_2^{(n)}(1)}$ of $H^{(n)}(0)$.

Now $H^{(n)}(1)$ may again contain hyperedges with cardinality 1: precisely those which have the form $e' = e \cap \mathbf{A}_2^{(n)}(1)$ for a hyperedge e of $H^{(n)}(0)$ which equals $e = \{a_1^n, \tilde{a}_1^n\}$ for some $a_1^n \in \mathbf{A}_1^{(n)}(1)$ and $\tilde{a}_1^n \in \mathbf{A}_2^{(n)}(1)$. Thus again eliminating those elements a_1^n from $\mathbf{A}_2^{(n)}(1)$ where $\{a_1^n\}$ is a hyperedge of $H^{(n)}(1)$, one arrives at a subhypergraph $H^{(n)}(2)$, and so on.

Formally, with $\mathbf{A}_2^{(n)}(0) := \mathbf{A}^n$, we set for $s \geq 1$

$$\begin{aligned}
\mathbf{A}_1^{(n)}(s) &:= \{a_1^n \in \mathbf{A}_2^{(n)}(s-1) : \{a_1^n\} \text{ is a hyperedge in } H^{(n)}(s-1)\},\\
\mathbf{A}_2^{(n)}(s) &:= \mathbf{A}^n \setminus \mathbf{A}_1^{(n)}(s),\\
H^{(n)}(s) &:= H^{(n)}(s-1)|_{\mathbf{A}_2^{(n)}(s)}.
\end{aligned}$$

After a finite number $S^{(n)}$ of steps we arrive at a hypergraph $H^{(n)} := H^{(n)}(S^{(n)})$ which is either empty or does not contain any hyperedge of cardinality 1. We denote the vertex set of $H^{(n)}$ by $\mathbf{A}_2^{(n)}$ and define $\mathbf{A}_1^{(n)} := \mathbf{A}^n \setminus \mathbf{A}_2^{(n)}$. Observe that

$$\begin{aligned}
&\mathbf{A}^n = \mathbf{A}_2^{(n)}(0) \supset \mathbf{A}_2^{(n)}(1) \supset \ldots \supset \mathbf{A}_2^{(n)}(S^{(n)}) = \mathbf{A}_2^{(n)},\\
&\mathbf{A}^{(n)}(1) \subset \ldots \subset \mathbf{A}^{(n)}(S^{(n)}) = \mathbf{A}_1^{(n)}.
\end{aligned} \tag{34}$$

The main step now is to prove $\mathbf{A}_1^{(n)} \subset (\mathbf{A}_1^{(1)})^n$ for every $n \geq 1$. Due to (34), this is implied by

$$\mathbf{A}_1^{(n)}(s) \subset (\mathbf{A}_1^{(1)})^n \qquad \text{for every } 1 \leq s \leq S^{(n)}. \tag{35}$$

For $n = 1$ nothing has to be proved. For every $n \geq 2$ we prove (35) by induction over s.

Let $n \geq 2$ and $s = 1$. If $a_1^n \in \mathbf{A}_1^{(n)}(1)$, then $\{a_1^n\}$ is a hyperedge in $H^{(n)}(0)$. As $H^{(n)}(0) = H^{(1)}(0)^n$, i. e. $H^{(n)}(0)$ is the n-fold square product of $H^{(1)}(0)$ with itself, this is only possible if $a_i \in \mathbf{A}_1^{(1)}(1) \subset \mathbf{A}_1^{(1)}$ for all $1 \leq i \leq n$.

Now assume (35) is proven for all $1 \leq \sigma \leq s$. Let $a_1^n \in \mathbf{A}_1^{(n)}(s+1)$, so that $\{a_1^n\}$ is a hyperedge in $H^{(n)}(s)$. This implies that there exists a hyperedge $e^{(n)} = \{a_1^n, a_{1,2}^n, \ldots, a_{1,\mu}^n\}$ in $H^{(n)}(0)$ such that for every $2 \leq \nu \leq \mu$ there exists a $1 \leq \sigma_\nu \leq s$ such that $a_{1,\nu}^n \in \mathbf{A}_1^{(n)}(\sigma_\nu)$. By the induction hypothesis, $a_{1,\nu}^n \in (\mathbf{A}_1^{(1)})^n$ for every ν.

Suppose $a_1^n \notin (\mathbf{A}_1^{(1)})^n$. Then $a_i \notin \mathbf{A}_1^{(1)}$ for some $1 \leq i \leq n$, so $a_i \in \mathbf{A}_2^{(1)}$. Hence for every hyperedge e of $H^{(1)}(0)$ containing a_i there is an $a_e \in \mathbf{A}_2^{(1)}$ not equal to a_i such that both a_i and a_e are contained in e.

Let $\{a_1^n, \tilde{a}_{1,2}^n, \ldots, \tilde{a}_{1,\tilde{\mu}}^n\}$ be any hyperedge in $H^{(n)}(0)$ containing a_1^n. As $H^{(n)}(0)$ is the n-fold square product of $H^{(1)}(0)$, there must be a $2 \leq \tilde{\nu} \leq \tilde{\mu}$ such that the i-th component of $\tilde{a}_{1,\tilde{\nu}}^n$ equals a_e for one of the hyperedges e of $H^{(1)}(0)$ containing a_i. In particular, $a_{1,\tilde{\nu}}^n \notin (\mathbf{A}_1^{(1)})^n$. However, this contradicts the existence of the hyperedge $e^{(n)} = \{a_1^n, a_{1,2}^n, \ldots, a_{1,\mu}^n\}$ in $H^{(n)}(0)$ which apart from a_1^n only contains elements of $(\mathbf{A}_1^{(1)})^n$.

This proves the claim (35), in particular $\mathbf{A}_1^{(n)} \subset (\mathbf{A}_1^{(1)})^n$. We therefore find that for every $n \in \mathbb{N}$, the number of messages that can be sent securely equals

$$N_{(\mathbf{T}_B, \mathbf{T}_C)}(n) = |\mathbf{A}^n| - |\mathbf{A}_1^{(n)}| \geq |\mathbf{A}|^n - |\mathbf{A}_1^{(1)}|^n.$$

If $\mathbf{A}_1^{(1)}$ is a strict subset of $\mathbf{A}$, then

$$C_0(\mathbf{T}_B, \mathbf{T}_C) = \lim_{n\to\infty} \frac{\log N_{(\mathbf{T}_B, \mathbf{T}_C)}(n)}{n} = \log|\mathbf{A}|.$$

Otherwise, $C_0(\mathbf{T}_B, \mathbf{T}_C)$ obviously equals 0. This proves Theorem 2.

4.5 Proof of Lemma 3

The proof of Lemma 3 is based on the fact that the labelling of the encoding sets of an M-code $\mathbf{F}$ is arbitrary. Let $\mathbf{F}$ be any (M, L, n)-code. k-fold concatenation of $\mathbf{F}$ with itself gives a $(M^k, L^{(k)}, kn)$-code $\mathbf{F}^k$. We show that the encoding sets of $\mathbf{F}^k$ can be labelled in such a way that

$$L^{(k)} = \frac{M^k - 1}{M - 1} L \tag{36}$$

is possible. The idea is to order the messages k-tuples $(m_1, \ldots, m_k)$ lexicographically. We define this recursively: For $k = 2$, the message pair (m_1, m_2) gets the label

$$l^{(2)}(m_1, m_2) = M(m_1 - 1) + m_2.$$

For $k \geq 2$ we set

$$l^{(k+1)}(m_1, \ldots, m_{k+1}) := M(l^{(k)}(m_1, \ldots, m_k) - 1) + m_{k+1}.$$

It is easy to check that the range of values of $l^{(k)}$ is $\{1, \ldots, M^k\}$.

For the concatenated code, we label the coding set $\mathbf{F}(m_1) \times \cdots \times \mathbf{F}(m_k)$ with $l^{(k)}(m_1, \ldots, m_k)$. Let $(c_1, \ldots, c_{kn})$ be an eavesdropper output sequence. As $\mathbf{F}$ is an (M, L, n)-code, for every $1 \leq i \leq n$ there are messages m_i, m_i' satisfying $m_i - m_i' \geq L - 1$ such that $(c_{(i-1)n+1}, \ldots, c_{in})$ is generated by both m_i and m_i'. It is easy to show by induction that the distance of $(m_1, \ldots, m_n)$ and $(m_1', \ldots, m_n')$ according to the labelling function $l^{(k)}$ is

$$l^{(k)}(m_1, \ldots, m_n) - l^{(k)}(m_1', \ldots, m_n') \geq \frac{M^k - 1}{M - 1} L.$$

Observe now that

$$\frac{L^{(k)} - 1}{M^k - 1} \longrightarrow \frac{L}{M - 1}$$

from below as $k \to \infty$. Thus every ratio $(L - 1)/(M - 1)$ can be improved by enlarging the blocklength, which proves the claim of Lemma 3.

References

[1] G. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1497–1510, June 2013.

[2] H. Li, L. Lai, and W. Zhang, "Communication requirement for reliable and secure state estimation and control in smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 476–486, Sept 2011.

[3] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2207–2229, Oct 1998.

[4] M. Fekete, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Math. Z.*, vol. 17, no. 1, pp. 228–249, 1923.

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

[6] M. Hellmuth, L. Ostermeier, and P. Stadler, "A survey on hypergraph products," *Math. Comput. Sci.*, vol. 6, no. 1, pp. 1–32, 2012.

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