Capacity Bounds for Identification With Effective Secrecy

Johannes Rosenberger,* Abdalla Ibrahim,* Boulat A. Bash,† Christian Deppe,* Roberto Ferrara,* Uzi Pereg‡
* TUM School of Computation, Information and Technology, Technical University of Munich,
Email: {johannes.rosenberger , abdalla.m.ibrahim , christian.deppe , roberto.ferrara }@tum.de
† Electrical and Computer Engineering Department, University of Arizona, Email: boulat@arizona.edu
‡ Faculty of Electrical and Computer Engineering and Helen Diller Quantum Center,
Technion—Israel Institute of Technology, Email: uzipereg@technion.ac.il

***Abstract*—An upper bound to the identification capacity of discrete memoryless wiretap channels is derived under the requirement of semantic effective secrecy, combining semantic secrecy and stealth constraints. A previously established lower bound is improved by applying it to a prefix channel, formed by concatenating an auxiliary channel and the actual channel. The bounds are tight if the legitimate channel is more capable than the eavesdropper's channel. An illustrative example is provided for a wiretap channel that is composed of a point-to-point channel, and a parallel, reversely degraded wiretap channel. A comparison with results for message transmission and for identification with only secrecy constraint is provided.**

I. Introduction

An increasing need for task-oriented and *semantic* communication paradigms that cater for a variety of tasks with different reliability, robustness, secrecy and privacy requirements can be observed across modern cyber-physical systems**Guenduez2023semantic**, [1], [2]. Pioneering work by Shannon [3] emphasized the problem of transmitting messages, where a decoder is expected to decide which message has been sent over a noisy channel among exponentially many possible hypotheses.

Identification (ID) [4]–[6] is a communication task where the receiver, Bob, selects one of many possible messages and tests whether this particular message was sent by the transmitter, Alice or not. It is assumed that Alice does not know Bob's chosen message; otherwise she could simply answer with "Yes" or "No", by sending a single bit. Given the nature of this very specific task, ID is in stark contrast to the conventional and general task of uniquely decoding messages, i.e. estimating which message was sent. Decoding is general in the sense that Bob can estimate any

function of Alice's message, while in ID, he can only compute whether it is the one of his interest or not, hence the function that Bob interested in is a simple indicator function. However, while the code sizes may only grow exponentially in the blocklength for the message-transmission task [3]–[6], a *doubly exponential* growth can be achieved in ID, if randomized encoding is used [6]. Randomized encoding allows the number of messages to be restricted not by number of possible distinct codewords, but rather by the number of input distributions to the channel that are pairwise distinguishable at the output [6], [7]. ID can have applications in various domains that span authentication tasks such as watermarking [8], [9], sensor communication [10], vehicle-to-X communication [11]–[13], among others.

From a more theoretical point of view, there are as well several interesting connections between ID and common randomness generation [14], as well as resolvability [7], [15]–[17] and soft-covering [18, p. 656], [19]–[21].

These connections lead to remarkable behavior of ID capacities in relation with security constraints. For example, in the semantic

secrecy regime, one can achieve the same ID rate, as if there were no security requirement, provided that the secrecy-capacity of the channel is positive [22], [23]. Stealth [24]–[27] requires that the adversary is prevented from determining whether Alice and Bob are communicating. As such, it may seem to be an even stronger requirement than secrecy. In general, however, neither stealth nor secrecy implies the other [25]. In contrast to covert communication[28]–[30], stealthy signals are not necessarily limited in power, but need to simulate a default distribution at the attacker's output, so that the attacker is unable to distinguish valuable information from noise. Thus, covert communication belongs to the same family of security problems as stealthy communication [25], [31].

The covert ID capacity for binary-input discrete memoryless wire-tap channels was determined by Zhang and Tan [32], and an achievable rate for ID under semantic effective secrecy has been established [33].

Here, we derive new bounds on the ID capacity of discrete memoryless wiretap channels [34], [35] under the semantic effective secrecy constraint. We improve the achievable rate from [33] by observing that the encoder can always prepend an auxiliary channel to the actual channel given to him. This may increase the achievable rate, e.g., if the channel is not more capable, i.e. $I(X;Y) \leq I(X;Z)$ for the input distribution P_X achieving the capacity without security constraints, the legitimate channel output Y, and attacker's output Z, but there exists an auxiliary random variable U such that $U - X - YZ$ is a Markov chain and $I(U;Y) \geq I(U;Z)$. Furthermore, we derive a converse bound, which is tight if the legitimate channel is more capable than the eavesdropper's one. We demonstrate

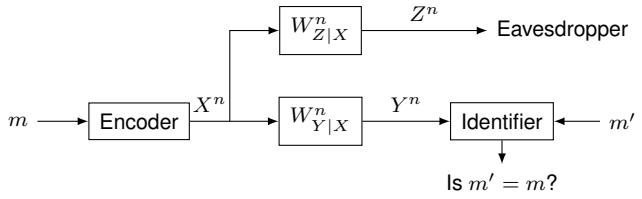

Figure 1. A general identification scheme in the presence of an eavesdropper, where $W^n_{YZ|X} = \prod_{i=1}^n W_{YZ|X}$ is a discrete memoryless wiretap channel. In contrast to conventional message transmission, the receiver does not decode the message m from the channel output Y^n, but chooses an m', and performs a statistical hypothesis test to decide whether m' equals m or not. In the effective secrecy setting, the eavesdropper wants to find out whether unexpected communication takes place or not, compared to some expected default behavior, and to identify whether an own message m'' equals m or not.

the results for a reversely degraded binary erasure broadcast channel, where each symbol consists of two bits, and for the first bit, the legitimate channel is stronger, and for the second bit, the eavesdropper's channel is better. Based on this example, we discuss the relation of the derived bounds, and the relations to capacities for other communication problems and constraints. Finally, we discuss future steps needed to obtain a generally tight converse bound.

The ID capacity of the discrete memoryless wiretap channel exhibits a similar dichotomy for semantic effective secrecy as for only semantic secrecy [22], but with a more stringent positivity condition and constraint. This is because for secrecy, only a small part of the ID codeword has to be secret [22], while for effective secrecy, the whole codeword must be stealthy [33].

This work is organized as follows. Section II sets the notation, channel model and communication task. In Section III, we review previous results and present our main theorems. We present the example and discuss the bounds in Section IV. In Section V, we develop an auxiliary ID converse, which permits convex constraints on the encoding distributions. It is then used in Section VI to prove the upper rate bound for effectively secrete ID. Finally, Section VII summarizes the results and discusses futher steps.

An extended version of this paper can be found on arXiv.

II. Preliminaries

The indicator function $\mathbb{1}(\cdot)$ evaluates to 1 if its argument is true, and to 0 if it is false.

The function $\log$ denotes the natural logarithm.

A. Channels

A *channel* with domain $\mathcal{X}$ and codomain $\mathcal{Y}$ is a conditional PMF $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. A *discrete memoryless channel* W (DMC) is a sequence $(W^n)_{n\in\mathbb{N}}$, where an input sequence x^n of block length $n \in \mathbb{N}$ is mapped to an an output sequence y^n with probability

$W^n(y^n|x^n) = \prod_{i=1}^n W(y_i|x_i)$.

A *wiretap channel* is a channel $W_{YZ|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y} \times \mathcal{Z})$, where we assume that for an input x, a legitimate receiver has access to the channel output $Y \sim W_{Y|X=x}$ and a passive adversary has access to to the output $Z \sim W_{Z|X=x}$, where the output distributions of the marginal channels $W_{Y|X}$ and $W_{Z|X}$ given input x are marginalizations of the joint output distribution $W_{YZ|X=x}$. A *discrete memoryless wiretap channel* $W_{YZ|X}$ (DMWC) is a sequence $(W^n_{YZ|X})_{n\in\mathbb{N}}$. A wiretap channel is *(stochastically) degraded* towards Y if there exists a channel $W_{Z|Y}$ such that $W_{Z|X} = W_{Y|X}W_{Z|Y}$.

The *Kullback-Leibler divergence* (KL divergence) between two PMFs $P, Q \in \mathcal{P}(\mathcal{X})$, where $Q(x) > 0$ if $P(x) > 0$, is defined by

$D(P\|Q) = \sum_{x\,:\,P(x)>0} P(x) \log \frac{P(x)}{Q(x)}$,

and two conditional PMFs $W, V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, the conditional KL divergence is defined by $D(W\|V|P) = \mathbb{E}_P[D(W(\cdot|X)\|V(\cdot|X))]$. The Shannon entropies and the mutual information are defined as usual. Note that for $(X, Y) \sim P \times W$ and any $Q \in \mathcal{P}(\mathcal{Y})$ such that the following expression is defined, it holds that

$$D(W\|Q|P) = I(X;Y) + D(PW\|Q). \tag{1}$$

B. Identification codes and effective secrecy

The task of ID with effective secrecy is described as follows: Consider a wiretap channel $W_{YZ|X}$ where Alice transmits a signal $X \in \mathcal{X}$, Bob, the legitimate receiver, receives $Y \in \mathcal{Y}$ and Willie, the adversary, receives $Z \in \mathcal{Z}$. Alice encodes a message $m \in [M] = \{1, \dots, M\}$ into X such that (a) Bob can test reliably whether Alice sent m' or not, for every $m' \in [M]$ *(identification)* and (b) Willie cannot distinguish whether Alice sent something sensible or nonsense *(stealth)*, nor can he identify whether Alice sent any particular $\tilde{m} \in [M]$ *(secrecy)*. The combination of stealth and secrecy is called *effective secrecy*. To this end, it is assumed that Willie will always classify the received signal as suspicious if it differs significantly from a prescribed distribution Q_Z. Let us formally define the involved codes and capacities:

An *M-code* for a discrete channel $W_{Y|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is a family $\{(E_m, \mathcal{D}_m)\}_{m=1}^M$ of encoding distributions $E_m \in \mathcal{P}(\mathcal{X})$ and decision sets $\mathcal{D}_m \subseteq \mathcal{Y}$. An (M, n)-code for a DMC W is an M-code for the channel $W_n = W^n$, i.e. $E_m \in \mathcal{P}(\mathcal{X}^n)$ and $\mathcal{D}_m \subseteq \mathcal{Y}^n$.

A *$(M|\lambda_1, \lambda_2)$-identification code* (ID code) is an M-code satisifying the conditions

$$\min_m E_m W_{Y|X}(\mathcal{D}_m) \geq 1 - \lambda_1, \tag{2}$$

$$\max_{m,m'\,:\,m\neq m'} E_m W_{Y|X}(\mathcal{D}_{m'}) \leq \lambda_2. \tag{3}$$

An *$(M|\lambda_1, \lambda_2, \delta, Q_Z)$ (semantically) effectively secret ID code* (ESID code) for a discrete wiretap channel $W_{YZ|X}$ is an $(M|\lambda_1, \lambda_2)$-ID-code where every encoding distribution E_m simulates Q_Z with precision $\delta > 0$ over $W_{Z|X}$, i.e.

$$\max_m D(E_m W_{Z|X}\|Q_Z) \leq \delta. \tag{4}$$

Semantic secrecy means that the constraint must hold for every message, not only for a particular random message distribution. Similarly, an $(M, n|\lambda_1, \lambda_2)$ ID code and an

$(M, n|\lambda_1, \lambda_2, \delta, Q_{Z^n})$ ESID code are defined for DMCs with block length n.

The *rate* of an (M, n) ID code is defined as $R = \frac{1}{n} \log \log M$. A rate R is Q_Z-ESID *achievable* over a wiretap channel $W_{YZ|X}$ if, for all $\lambda_1, \lambda_2, \delta > 0$ and sufficiently large n, there exists an $(2^{2^{nR}}, n|\lambda_1, \lambda_2, \delta, Q_Z)$ ESID code. The Q_{Z^n}-ESID capacity $\mathsf{C}_{\mathsf{ESID}}(W_{YZ|X}, Q_{Z^n})$ is the supremal rate that is Q_{Z^n}-ESID achievable over $W_{YZ|X}$.

III. Results

Consider the prior result from [33].

Proposition 1 ([33, Theorem 1])**.** *For any $Q_Z \in \mathcal{P}(\mathcal{Z})$, the Q_Z^n-ESID capacity of a discrete memoryless wiretap channel $W_{YZ|X}$ satisfies*

$$\mathsf{C}_{ESID}(W_{YZ|X}, Q_Z^n) \geq \max_{\substack{P_X \in \mathcal{P}(\mathcal{X}) \\ P_X W_{Z|X} = Q_Z \\ I(X;Y) \geq I(X;Z)}} I(X;Y). \tag{5}$$

In Proposition 1, the constraint $I(X;Y) \geq I(X;Z)$ can be relaxed to $I(U;Y) \geq I(U;Z)$, for any auxiliary random variable U having finite support and satisfying the Markov condition $U - X - YZ$, by applying Proposition 1 to the virtual channel $P_{YZ|U} = P_{X|U} W_{YZ|X}$:

Corollary 1. *The Q_Z^n-ESID capacity of a DMWC $W_{YZ|X}$ satisfies*

$$\mathsf{C}_{ESID}(W_{YZ|X}, Q_Z^n) \geq \max_{\substack{P_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}) \\ P_X W_{Z|X} = Q_Z \\ I(U;Y) \geq I(U;Z)}} I(U;Y), \tag{6}$$

where $\mathcal{U}$ is any finite set.

On the other hand, we prove the following upper bound.

Theorem 1. *The Q_Z^n-ESID-capacity of a DMWC $W_{YZ|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y} \times \mathcal{Z})$ is 0 if $I(U;Y) < I(U;Z)$ or $P_X W_{Z|X} \neq Q_Z$, for all P_{UX} such that $U - X - YZ$ forms a Markov chain. Otherwise, it satisfies*

$$\mathsf{C}_{ESID}(W_{YZ|X}, Q_Z^n) \leq \max_{\substack{P_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}) \\ P_X W_{Z|X} = Q_Z \\ I(U;Y) \geq I(U;Z)}} I(X;Y), \tag{7}$$

where $|\mathcal{U}| \leq |\mathcal{X}| + 2$. If $W_{Y|X}$ is more capable than $W_{Z|X}$,

$$\mathsf{C}_{ESID}(W_{YZ|X}, Q_Z^n) = \max_{\substack{P_X \in \mathcal{P}(\mathcal{X}) \\ P_X W_{Z|X} = Q_Z}} I(X;Y). \tag{8}$$

The proof follows in Section VI.

Remark 1. The lower bound in Corollary 1 and the upper bound in Theorem 1 coincide only for channels where the optimal capacity is achieved with $U = X$. In the following section, we demonstrate the gap at the example of a reversely degraded wiretap channel, where $U \neq X$ is optimal.

Remark 2. If the Q_Z^n-ESID capacity is zero, then effectively secret communication with any positive rate impossible. Yet, this does not necessarily imply that communication is impossible. It can simply mean that the code size grows slower than doubly-exponentially in the block length, since we defined the rate as $R = \frac{1}{n} \log \log M$.

For example, for covert codes $O(\sqrt{n})$ bits can be sent in n channel uses. However, Ahlswede [36, Lemmas 89, 90 and Remark 92] proved that, if for sufficiently small $\lambda_1, \lambda_2, \delta$ and sufficiently large n, the secrecy condition $I(U;Y) \geq I(U;Z)$ is violated for all U, then secret communication is impossible, hence also effectively secret communication.

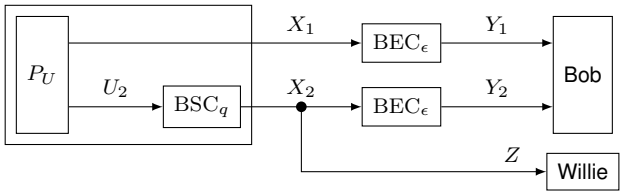

Figure 2. Structure of the product of two reversely degraded broadcast example

IV. Example: Reversely degraded broadcast channels

To demonstrate the relation of the capacity bounds in Proposition 1, Corollary 1, and Theorem 1 and contrast it with message transmission, we consider two reversely degraded binary broadcast channels used in parallel, as shown in Figure 2. In this section, in a binary setting, we let $\log := \log_2$. Every input symbol consists of two bits, i.e. $\mathcal{X} = \{0,1\}^2$. The bit X_1 is sent only to Bob, over a binary erasure channel, BEC_ϵ, described by the transition matrix

$\mathrm{BEC}_\epsilon = \left(\begin{smallmatrix} 1-\epsilon & \epsilon & 0 \\ 0 & \epsilon & 1-\epsilon \end{smallmatrix}\right)$,

where $\frac{1}{2} < \epsilon \leq 1$ is the erasure probability, the output alphabet is $\{0, e, 1\}$, and e denotes an erasure. The bit X_2 is also sent to Bob over a $\mathrm{BEC}(\epsilon)$ with the same erasure probability, while Willie observes it noiselessly. We have thereby the assignments $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$, $Z = X_2$, $Y_1 \sim \mathrm{BEC}_\epsilon(\cdot|X_1)$, and $Y_2 \sim \mathrm{BEC}_\epsilon(\cdot|X_2)$. The output alphabets of the described channel are $\mathcal{Y} = \{0, e, 1\}^2$ and $\mathcal{Z} = \mathcal{X}$. This channel belongs to the class of reversely degraded broadcast channels [37, Page 127], i.e. $P_{Y,Z|X} = P_{Y_1,Y_2,Z|X_1,X_2} = P_{Y_1|X_1} \cdot P_{Z|X_2} \cdot P_{Y_2|Z}$.

Let $P_{X_1} = (p_{X_1}, 1 - p_{X_1})$, i.e. $P_{X_1}(0) = p_{X_1}$, and similarly $P_{X_2} = (p_{X_2}, 1 - p_{X_2})$. Since X_1 is perfectly secret, Alice chooses $p_{X_1} = \frac{1}{2}$, which maximizes the mutual information $I(X_1; Y_1) = (1 - \epsilon) H_2(p_{X_1}) = 1 - \epsilon$, by [38, Eq. (7.15)]. Suppose the default distribution to simulate for effective secrecy is $Q_Z = (\frac{1}{2}, \frac{1}{2})$. Then, (6) and (7) require that $p_{X_2} = \frac{1}{2}$. Thus, the mutual informations of the marginal channels are

$$\begin{aligned} I(X;Y) &= I(X_1;Y_1) + I(X_2;Y_2) && (9) \\ &= (1-\epsilon)(H_2(p_{X_1}) + H_2(p_{X_2})) && (10) \\ &= 2(1-\epsilon), && (11) \\ I(X;Z) &= H_2(p_{X_2}) = 1, && (12) \end{aligned}$$

where $H_2(p) = -p \log p - (1-p) \log(1-p)$. Since $\epsilon > \frac{1}{2}$, $P_{Y|X}$ is less capable than $P_{Z|X}$, i.e.

$I(X;Y) < I(X;Z)$,

and Proposition 1 guarantees no achievable Q_Z^n-ESID rate. Yet, Alice can achieve Q_Z^n-ESID, by letting $p_{U_2} = \frac{1}{2}$

$P_{U_2} = (p_{U_2}, 1-p_{U_2})$, $X_2 \sim \mathrm{BSC}_q(\cdot|U_2)$, and $U = (X_1, U_2)$,

where $\mathrm{BSC}_q = \left(\begin{smallmatrix} 1-q & q \\ q & 1-q \end{smallmatrix}\right)$ is a binary symmetric channel with crossover probability q. Elementary calculations show that

$$p_{X_2} = p_{U_2}(1-q) + (1-p_{U_2})q = \frac{1}{2}, \tag{13}$$

$$I(U;Y) = I(X_1;Y_1) + I(U_2;Y_2) \tag{14}$$

$$= (1-\epsilon)(2 - H_2(q)), \tag{15}$$

$$I(U;Z) = I(U_2;Z) = 1 - H_2(q), \tag{16}$$

and

$I(U;Y) \geq I(U;Z)$, for all $\epsilon \leq 1/(2-H_2(q))$.

Hence, by Corollary 1 and Theorem 1,

$$2(1-\epsilon) \geq I(X;Y) \geq \mathsf{C}_{\mathsf{ESID}}(W_{YZ|X}, Q_Z^n) \tag{17}$$

$$\geq I(U;Y) \tag{18}$$

$$= (1-\epsilon)(2 - H_2(q)), \tag{19}$$

where the gap is given by

$I(X;Y) - I(U;Y) = (1-\epsilon)H_2(q)$.

For comparison, given $P_{X_2|U_2} = \mathrm{BSC}_q$, an upper rate bound for effectively secret message transmission is [25, Theorem 1.1]

$$R_{\mathsf{EST}} \leq I(U;Y) - I(U;Z) \tag{20}$$

$$\leq 1 - \epsilon \tag{21}$$

$$\leq I(X_1;Y). \tag{22}$$

On the other hand, the ID capacity with only secrecy, without stealth, is given by $I(X;Y)$, since there exists P_U such that $I(U;Y) \geq I(U;Z)$ This suggests that for transmission with effective secrecy, it is optimal for Alice to only encode into the first bit, X_1, while effectively secret ID codes can increase the rate by exploiting both bits (X_1, U_2). Figure 3 displays the mutual informations for varying $0 \leq p_{U_2} = p_{X_2} \leq 1$, $q = \frac{1}{8}$, and hence $\epsilon = 1/(2-H_2(q)) = \frac{3}{8} - \frac{7}{8}\log_2(\frac{7}{8}) \approx 0.6866$. Thus, for $p_{U_2} = \frac{1}{2}$, we have $I(U;Y) = I(U;Z)$.

V. Auxiliary identification converse

Consider the hypothesis testing divergence

$$D_\alpha(P\|Q) := \sup\left\{\gamma : P\left(\log\frac{P(Y)}{Q(Y)} \leq \gamma\right) \leq \alpha\right\}, \tag{23}$$

where $Y \sim P$, and let $(P \otimes Q)(x,y) := P(x)Q(y)$.

Lemma 1. *Let $\lambda_1, \lambda_2, \eta > 0$ and $\alpha := \lambda_1 + \lambda_2 + 2\eta < 1$. For every $(M|\lambda_1, \lambda_2)$ ID code $\{(E_m, \mathcal{D}_m)\}_{m=1}^M$ for a channel $W_{Y|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, its size is bounded by*

$$\log\log M \leq \max_{P_X \in \mathcal{E}} \min_{Q \in \mathcal{P}(\mathcal{Y})} D_\alpha(P_{XY}\|P_X \otimes Q) + \epsilon \tag{24}$$

$$\leq \max_{P_X \in \mathcal{E}} \frac{1}{1-\alpha} I(X;Y) + \epsilon, \tag{25}$$

where $P_{XY} = P_X \times W_{Y|X}$, $\epsilon = \log\log|\mathcal{X}| + 3\log(1/\eta) + 2$, and $\mathcal{E} = \{P_V E_{m=V} : P_V \in \mathcal{P}([M])\}$ is the convex hull of all encoding distributions.

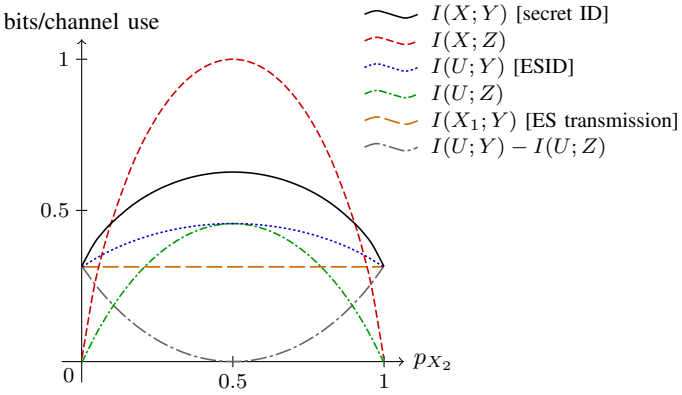

Figure 3. Mutual informations for the reversely degraded wiretap channel in Figure 2, and communication tasks where the mutual information is achievable, where $q = \frac{1}{8}$, $\epsilon = 0.6866 \approx 1/(2-H_2(q))$, and $P_{X_1} = Q_Z = (\frac{1}{2}, \frac{1}{2})$. For secret ID and effectively secret message transmission, the given bounds are tight, if $p_{U_2} = \mathrm{BSC}(q)$, for any q.

Proof. By [17, Corollary 2 and Lemma 1], we have that

$$\log\log M \leq \max_{P_X \in \mathcal{P}(\mathcal{X})} \min_{Q \in \mathcal{P}(\mathcal{Y})} D_\alpha(P_{XY}\|P_X \otimes Q) + \epsilon. \tag{26}$$

The maximization over $\mathcal{P}(\mathcal{X})$ is introduced in the proof of [17, Theorem 1] to establish the upper bound

$$\frac{1}{2}\left[(E_m \times W_{Y|X})(\mathcal{S}) + (E_{m'} \times W_{Y|X})(\mathcal{S})\right] \leq \sup_{P_X \in \mathcal{P}(\mathcal{X})} P_X \times W_{Y|X}(\mathcal{S}), \tag{27}$$

for some set $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Y}$, which ultimately leads to the maximization in (26). Clearly, a maximization over all E_m, $m \in [M]$ would suffice in (27).

To establish the minimax equality in [17, Corollary 2], Watanabe used the fact that the supremum is taken over a compact and convex set. Since no other properties of $\mathcal{P}(\mathcal{X})$ are used in [17], it suffices to maximize over the convex hull $\mathcal{E}$ of all encoding distributions,

as in (24). By Markov's inequality,

$$D_\alpha(P\|Q) = \sup\left\{\gamma : P\left(\log\frac{P(Y)}{Q(Y)} > \gamma\right) \geq 1-\alpha\right\} \tag{28}$$

$$\leq \inf\left\{\gamma : P\left(\log\frac{P(Y)}{Q(Y)} \geq \gamma\right) \leq 1-\alpha\right\} \tag{29}$$

$$\leq \frac{1}{1-\alpha}\mathbb{E}\left[\log\frac{P(Y)}{Q(Y)}\right] \tag{30}$$

$$\leq \frac{1}{1-\alpha} D(P\|Q), \tag{31}$$

where the first inequality holds since $p(\gamma) = P\left(\log\frac{P(Y)}{Q(Y)} \geq \gamma\right)$ is a decreasing function. Thus, for every $P_X \in \mathcal{P}(\mathcal{X})$,

$$\min_{Q \in \mathcal{P}(\mathcal{Y})} D_\alpha(P_X \times W_{Y|X}\|P_X \otimes Q) \leq \frac{1}{1-\alpha} D(P_X \times W_{Y|X}\|P_X \otimes P_X W_{Y|X}) \tag{32}$$

$$= \frac{1}{1-\alpha} I(X;Y), \tag{33}$$

and Lemma 1 follows. ■

VI. Proof of Theorem 1

Consider any $(M, n|\lambda_1, \lambda_2, \delta, Q_Z)$ ESID code $\{(E_m, \mathcal{D}_m)\}_{m=1}^M$ for a DMWC $W_{YZ|X} : \mathcal{X} \to \mathcal{P}(\mathcal{Y} \times \mathcal{Z})$, where $\lambda_1, \lambda_2, \eta > 0$ satisfy $\alpha := \lambda_1 + \lambda_2 + 2\eta < 1$. By Lemma 1, the rate is upper-bounded by

$$R = \frac{1}{n} \log\log M \leq \max_{P_{X^n} \in \mathcal{E}} \frac{1}{n(1-\alpha)} I(X^n; Y^n) + \frac{\epsilon}{n}, \quad (34)$$

where $\epsilon = \log\log |\mathcal{X}^n| + 3\log(1/\eta) + 2$, and $\mathcal{E}$ is the convex hull of all encoding distributions. By the chain rule,

$$\frac{1}{n} I(X^n; Y^n) = \frac{1}{n} \sum_{i=1}^n I(X^n; Y_i | Y^{i-1}) \quad (35)$$

$$= I(X^n; Y_T | T, Y^{T-1}) \quad (36)$$

$$\leq I(X_T; Y_T), \quad (37)$$

where $T \sim P_T(i) = \frac{1}{n}\mathbb{1}(1 \leq i \leq n)$, and (37) follows from the concavity of the mutual information in the input argument, and the Markov condition $Y^{i-1} - X^{i-1}X_{i+1}^n - X_i - Y_i$ for every $i \in [n]$.

In the following, we single-letterize the constraints on $\mathcal{E}$. Similarly to [25, Eq. (1.49)], for any $P_{Z^n} \in \mathcal{P}(\mathcal{Z}^n)$,

$$\delta = D(P_{Z^n} \| Q_Z^n) \quad (38)$$

$$= \sum_{z^n} P_{Z^n}(z^n) \sum_{i=1}^n \log \frac{1}{Q_Z(z_i)} - H(Z^n) \quad (39)$$

$$\geq \sum_{i=1}^n \sum_z P_{Z_i}(z) \log \frac{1}{Q_Z(z)} - \sum_{i=1}^n H(Z_i) \quad (40)$$

$$= n \sum_{i=1}^n P_T(i) D(P_{Z_i} \| Q_Z) \quad (41)$$

$$\geq n D(P_{Z_T} \| Q_Z). \quad (42)$$

By [36, Lemma 90], for sufficiently small $\lambda_1, \lambda_2, \delta > 0$,

$$\max_{P_A \in \mathcal{P}([M])} I(A; Y^n) \geq \delta \quad (43)$$

$$\geq \max_{m \in [M]} D(E_m W_{Z|X}^n \| Q_Z^n) \quad (44)$$

$$\geq \max_{P_A \in \mathcal{P}([M])} D(E_{m=A} W_{Z|X}^n \| Q_Z^n | P_A) \quad (45)$$

$$\geq \max_{P_A \in \mathcal{P}([M])} I(A; Z^n). \quad (46)$$

Thus, there exists $P_A \in \mathcal{P}([M])$ such that

$$0 \leq \frac{1}{n}[I(A; Y^n) - I(A; Z^n)] \quad (47)$$

$$= I(VA; Y_T | V) - I(VA; Z_T | V) \quad (48)$$

$$\leq \max_v \max_{P_{BX|V=v}} [I(B; Y_T | V = v) - I(B; Z_T | V = v)] \quad (49)$$

$$\leq \max_{P_{BX}} [I(B; Y) - I(B; Z)], \quad (50)$$

where $B = (V, A)$, $V = (T, Y_1, \ldots, Y_{T-1}, Z_{T+1}, \ldots, Z_n)$, the maximizations are with respect to $D(P_B P_{X_T|B} W_{Z|X} \| Q_Z) \leq \frac{\delta}{n}$, and (48) follows from [39, Lemma 17.12]. By [39, Lemmas 15.4 and 15.5], we can replace P_B by $P_U \in \mathcal{P}(\mathcal{U})$, $|\mathcal{U}| \leq |\mathcal{X}| + 2$, such that $I(U;Y) = I(B;Y)$, $I(U;Z) = I(B;Z)$, and $P_U P_{X|B} = P_B P_{X|B}$.

Since the mutual information is continuous and the set $\{P_{UX} : D(P_X W_{Z|X} \| Q_Z) \leq \delta\}$ is compact, for $\eta = e^{-\sqrt{n}}$, we have that

$$\mathsf{C}_{\mathsf{ESID}}(W_{YZ|X}, Q_Z^n) \quad (51)$$

$$\leq \inf_{\lambda_1, \lambda_2, \delta > 0} \lim_{n \to \infty} \Bigg[o(1) + \frac{1}{1 - \lambda_1 - \lambda_2 - o(1)} \max_{\substack{P_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}) \\ D(P_X W_{Z|X} \| Q_Z) \leq \delta/n \\ I(U;Y) \geq I(U;Z)}} I(X;Y) \Bigg] \quad (52)$$

$$= \max_{\substack{P_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}) \\ P_X W_{Z|X} = Q_Z \\ I(U;Y) \geq I(U;Z)}} I(X;Y). \quad (53)$$

For more capable channels, $I(X;Y) \geq I(X;Z)$, for all $P_X \in \mathcal{P}(\mathcal{X})$ and hence, the upper bound is achievable, by Proposition 1. This completes the proof of Theorem 1. ■

VII. Conclusion

In Corollary 1, we improved the lower bound on the Q_Z^n-ESID capacity in the case $I(X;Y) < I(X;Z)$, where Q_Z^n is a product distribution and P_X satisfies $P_X W_{Z|X} = Q_Z$. In Theorem 1, we complement this result by an upper bound that is tight if $I(X;Y) \geq I(X;Z)$. The example in Section IV illustrates that in case $I(X;Y) < I(X;Z)$, the achievability gap between Corollary 1 and Theorem 1 can be substantial, as is the rate advantage of ID compared to message transmission. It seems likely that the lower bound in Corollary 1 is tight, by results from resolvability theory [40], since the whole codeword is subject to the stealth constraint. The difficult part in finding a more stringent converse bound seems to be the introduction of an auxiliary channel, as demonstrated in Section IV, where the number of possible input sequences is suitably bounded (see Lemma 1 and the discussion of the gap in Ahlswede's broadcast converse in [41, Section 2.4]). This is a non-trivial task for ID, and the authors are not aware of any ID capacity result involving auxiliary variables in the rate bound. Usually, in converse proofs [37], the message is obtained as an auxiliary variable from Fano's inequality, and then is single-letterized. However, the mutual information between the message and the channel output cannot be an upper bound to the ID capacity, since ID codes transmit mainly randomness, and the capacity of memoryless channels is achieved with codes, where only few ($\sqrt{n}$) codeword symbols depend on the message at all [42]. To close the achievability gap, new methods need to be developed to introduce auxiliary variables in ID converse bounds. This would also be a crucial step in the development of further multi-user converses for ID and many other communication tasks[36], [43], e.g. for the broadcast channel [41], [43]. To this end, observe that any suitably bounded auxiliary channel forms a polytope with extremal points $P_{X|U=u}$, $u \in \mathcal{U}$, where $\mathcal{U}$ is the auxiliary alphabet. This polytope must include the set

of stealthy encoding distributions. If such a pre-channel exists, the usual ID converse can be applied to it, to obtain an upper bound that matches the bound in Corollary 1.

ACKNOWLEDGEMENT

The authors thank Constantin Runge (Technical University of Munich) for helpful discussions.

J. Rosenberger, A. Ibrahim and C. Deppe acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the program of "Souverän. Digital. Vernetzt." Joint project 6G-life, project identification number: 16KISK002. C. Deppe and R. Ferrara were further supported in part by the BMBF within the grant 16KIS1005. C. Deppe was also supported by the DFG within the project DE1915/2-1. B. Bash acknowledges the support from the US National Science Foundation under Grant CCF-2006679. U. Pereg acknowledges the financial support of the Israel VATAT Junior Faculty Program in Quantum Science and Technology and the German-Israeli Project Cooperation (DIP).

REFERENCES

[1] J. Cabrera, H. Boche, C. Deppe, R. F. Schaefer, C. Scheunert, and F. H. P. Fitzek, "6G and the Post-Shannon-Theory," in *Shaping Future 6G Networks: Needs, Impacts and Technologies*, E. Bertin, N. Crespi, and T. Magedanz, Eds., Wiley-Blackwell, Nov. 2021, ISBN: 978-1-119-76551-6. DOI: https://doi.org/10.1002/9781119765554.ch16.

[2] S. Rezwan, J. A. Cabrera, and F. H. P. Fitzek, "Network functional compression for control applications," in *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2022, pp. 1–6. DOI: 10.1109/ICECCME55909.2022.9988100.

[3] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Techn. J.*, vol. 27, no. 4, pp. 623–656, Oct. 1948. DOI: 10.1002/j.1538-7305.1948.tb00917.x.

[4] A. C.-C. Yao, "Some Complexity Questions Related to Distributive Computing (Preliminary Report)," in *Proc. 11th Ann. ACM Symp. Th. Comp.*, ser. STOC '79, Atlanta, Georgia, USA: Association for Computing Machinery, 1979, pp. 209–213, ISBN: 9781450374385. DOI: 10.1145/800135.804414.

[5] J. JaJa, "Identification is easier than decoding," in *26th Ann. Symp. Foundations Comp. Sci. (sfcs 1985)*, Oct. 1985, pp. 43–50. DOI: 10.1109/SFCS.1985.32.

[6] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 15–29, 1989. DOI: 10.1109/18.42172.

[7] T. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993. DOI: 10.1109/18.256486.

[8] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1410–1422, 2001. DOI: 10.1109/18.923724.

[9] Y. Steinberg, "Watermarking identification for private and public users: the broadcast channel approach," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2002, pp. 5–7. DOI: 10.1109/ITW.2002.1115400.

[10] O. Günlü, J. Kliewer, R. F. Schaefer, and V. Sidorenko, "Doubly-Exponential Identification via Channels: Code Constructions and Bounds," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 1147–1152. DOI: 10.1109/ISIT45174.2021.9518193.

[11] H. Boche and C. Deppe, "Secure Identification for Wiretap Channels; Robustness, Super-Additivity and Continuity," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1641–1655, 2018. DOI: 10.1109/TIFS.2018.2797004.

[12] H. Boche and C. Arendt, *Communication method, mobile unit, interface unit, and communication system*, US Patent 10,959,088, 2021.

[13] J. Rosenberger, U. Pereg, and C. Deppe, "Identification over Compound MIMO Broadcast Channels," in *Proc. IEEE Int. Conf. Comm. (ICC'2022)*, 2022.

[14] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, 1998. DOI: 10.1109/18.651026.

[15] Y. Steinberg, "New converses in the theory of identification via channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 984–998, 1998. DOI: 10.1109/18.669139.

[16] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, 2006. DOI: 10.1109/TIT.2006.871040.

[17] S. Watanabe, "Minimax converse for identification via channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 1, pp. 25–34, Jan. 2022, ISSN: 1557-9654. DOI: 10.1109/TIT.2021.3120033.

[18] R. Ahlswede, *Combinatorial Methods and Models, Rudolf Ahlswede's Lectures on Information Theory 4* (Foundations in Signal Processing, Communications and Networking), A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Springer International Publishing, 2018. DOI: 10.1007/978-3-319-53139-7.

[19] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975. DOI: 10.1109/TIT.1975.1055346.

[20] P. Cuff, "A stronger soft-covering lemma and applications," in *2015 IEEE Conference on Communications and Network Security (CNS)*, 2015, pp. 40–43. DOI: 10.1109/CNS.2015.7346808.

[21] H. Boche and C. Deppe, "Secure Identification Under Passive Eavesdroppers and Active Jamming Attacks," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 472–485, 2019. DOI: 10.1109/TIFS.2018.2854729.

[22] R. Ahlswede and Z. Zhang, "New directions in the theory of identification via channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1040–1050, 1995. DOI: 10.1109/18.391249.

[23] W. Labidi, C. Deppe, and H. Boche, "Secure identification for gaussian channels," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 2872–2876.

[24] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, "Reliable, deniable and hidable communication: A quick survey," in *2014 IEEE Information Theory Workshop (ITW 2014)*, 2014, pp. 227–231. DOI: 10.1109/ITW.2014.6970826.

[25] J. Hou, G. Kramer, and M. Bloch, "Effective secrecy: Reliability, confusion, and stealth," in *Information Theoretic Security and Privacy of Information Systems*, R. F. Schaefer, H. Boche, A. Khisti, and H. V. Poor, Eds., Cambridge University Press, 2017, pp. 3–20. DOI: 10.1017/9781316450840.002.

[26] J. Song, Q. Zhang, S. Kadhe, M. Bakshi, and S. Jaggi, "Stealthy communication over adversarially jammed multipath networks," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7473–7484, 2020. DOI: 10.1109/TCOMM.2020.3022785.

[27] M. Bloch, O. Günlü, A. Yener, *et al.*, "An Overview of Information-Theoretic Security and Privacy: Metrics, Limits and Applications," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 5–22, Mar. 2021, ISSN: 2641-8770. DOI: 10.1109/JSAIT.2021.3062755.

[28] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013. DOI: 10.1109/JSAC.2013.130923.

[29] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: Fundamental limits of covert wireless communication," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 26–31, 2015. DOI: 10.1109/MCOM.2015.7355562.

[30] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, 2016. DOI: 10.1109/TIT.2016.2530089.

[31] D. Lentner and G. Kramer, "Stealth communication with vanishing power over binary symmetric channels," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 822–827. DOI: 10.1109/ISIT44484.2020.9174381.

[32] Q. Zhang and V. Y. F. Tan, "Covert Identification Over Binary-Input Discrete Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5387–5403, Aug. 2021. DOI: 10.1109/tit.2021.3089245.

[33] A. Ibrahim, R. Ferrara, and C. Deppe, "Identification under Effective Secrecy," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2021, pp. 1–6. DOI: 10.1109/ITW48936.2021.9611488.

[34] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975. DOI: 10.1002/j.1538-7305.1975.tb02040.x.

[35] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[36] R. Ahlswede, *Identication and Other Probabilistic Models, Rudolf Ahlswede's Lectures on Information Theory 6* (Foundations in Signal Processing, Communications and Networking), 1st Edition, A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Cham: Springer International Publishing, 2021, ISBN: 978-3-030-65072-8. DOI: 10.1007/978-3-030-65072-8.

[37] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011. DOI: 10.1017/CBO9781139030687.

[38] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Ltd, Apr. 2005, ISBN: 9780471748823. DOI: 10.1002/047174882x.ch15.

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

[40] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *2013 13th Canad. Workshop Inf. Th.*, 2013, pp. 76–81. DOI: 10.1109/CWIT.2013.6621596.

[41] A. Bracher, "Identification and Zero-Error Codes," Ph.D. dissertation, ETH Zurich, Konstanz, 2016, ISBN: 3-86628-574-4. DOI: 10.3929/ethz-a-010739015.

[42] R. Ahlswede and G. Dueck, "Identification in the presence of feedback—a discovery of new capacity formulas," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 30–36, 1989. DOI: 10.1109/18.42173.

[43] R. Ahlswede, "General theory of information transfer: Updated," *Discr. Appl. Math.*, vol. 156, no. 9, pp. 1348–1388, May 2008. DOI: 10.1016/j.dam.2007.07.007.