Achievable Error Exponents for the Private Fingerprinting Game

Anelia Somekh-Baruch and Neri Merhav
Department of Electrical Engineering
Technion - I.I.T., Haifa 32000, Israel.
{anelia@tx, merhav@ee}.technion.ac.il

Abstract

Fingerprinting systems in the presence of *collusive* attacks are analyzed as a game between a fingerprinter and a decoder, on the one hand, and a coalition of two or more attackers, on the other hand. The fingerprinter distributes, to different users, different fingerprinted copies of a host data (*covertext*), drawn from a memoryless stationary source, embedded with different fingerprints. The coalition members create a forgery of the data while aiming at erasing the fingerprints in order not to be detected. Their action is modelled by a multiple access channel (MAC). We analyze the performance of two classes of decoders, associated with different kinds of error events. The decoder of the first class aims at detecting the *entire* coalition, whereas the second is satisfied with the detection of *at least one* member of the coalition. Both decoders have access to the original covertext data and observe the forgery in order to identify member/s of the coalition. Motivated by a worst-case approach, we assume that the coalition of attackers is informed of the hiding strategy taken by the fingerprinter and the decoder, while they are uninformed of the attacking scheme. Single letter expressions for the error exponents of the two kinds are obtained, a decoder that is optimal with respect to the two kinds of errors is introduced, and the worst-case attack channel is characterized.

1 Introduction

In fingerprinting systems several copies of the same host data are embedded with different fingerprints (that designate, e.g., the different digital signatures or serial numbers of the copies they are provided with) and distributed to different users . The fingerprints identify one of many users in order to enable copyright protection. In this situation, two or more users can form a coalition, and collusive attacks on the fingerprinting system are possible and have to be taken into account in the code design. Each of the coalition members contributes his distinct fingerprinted copy in order to create a better forgery. Hence, the fingerprinting problem can be thought of as a game between the fingerprinter and the coalition of attackers.

As mentioned in [32], the fingerprinting game is closely related to (and is actually an extension of) the watermarking game, that in turn can be modelled as a coded communication system equipped with side information, for a single user as opposed to one of many users. Watermarking systems have been studied from the information theoretic point of view in several works[1] (see e.g., [4], [7], [8], [14], [19], [21], [23], [24], [25], [30], [31] and [33]). Several researchers (see, e.g., [1]-[3], [5], [6], [9], [12], [15], [17], [18], [26], [27], [29], [34]-[36] and references therein) have proposed and analyzed fingerprinting systems that aim at protection against collusion attacks under various conditions. The research in this problem area has largely focused on combinatorial analysis and algorithmic issues, whereas there has been much less work on information–theoretic aspects. As an exception to the last statement, we mention several papers, such as [2], [20], [32] and [34].

In [32], we have presented and analyzed a game-theoretic model of private[2] fingerprinting systems in the presence of colluding attacks. The players of the game are on the one hand an encoder-decoder and on the other hand, a few attackers. The decoder is facing the rather complicated goal of reliably identifying coalition members based on the forgery and the covertext. The distortion between the forgery and the original data should not exceed a certain level. A realistic worst-case approach, taken in [32], is based on the assumption that the attackers are informed of the covertext distribution and the coding-decoding strategy (up to a random secret key), whereas the encoder and decoder are not informed of the attack strategy. Thus, the encoder and decoder are assumed to adopt a random coding strategy as a means of protection against malicious attackers. Random coding is enabled by a secret key that specifies the particular codebook that has been drawn, which is shared by the encoder and decoder. The action of the attackers is allowed to be stochastic, and therefore one can model the attack by a MAC, whose output is the forgery and whose inputs are fingerprinted copies that are observed by the coalition members. The problem of the maximin game between the fingerprinter and the MAC applied by the users is addressed in [32]. Two types of decoders are considered: a *single-output (SO) decoder* whose single output is a message index and a *multiple-output (MO) decoder* whose output is a list containing L message indices (where L is the size of the coalition). The two decoders aim at detecting *only one* member of the coalition. When the SO decoder is concerned, an error is declared if its output does not belong to the coalition, whereas when the MO decoder is concerned, an error is declared when none of its outputs belongs to the coalition. The reason for aiming at this goal (adopted also in, e.g., [2],[29] and references therein) is that when the forgery is required to resemble to *at least* one of the fingerprinted copies observed by the attacker target of detecting the entire coalition is impossible since the attacker can decide to almost ignore one of its inputs, and thus prevent reliable decoding of the input that was ignored. It is assumed in [32] that the encoder uses constant composition (CC) codes and under this assumption, a single-letter expression for the capacity of the private fingerprinting game with respect to (w.r.t.) the two types of decoders is found, and it is shown that their capacities

[1] For a more comprehensive survey, see [32].

[2] Two models of the game may be considered: the private game in which the covertext is available also at the decoder's side, and the public game, where it is only available to the encoder.

are the same. Asymptotically optimal strategies, taken by the adversaries, are characterized. Also, lower bounds on the error exponents of the two types of decoders are derived.

In this paper, we analyze private fingerprinting systems from a different perspective. Two kinds of error events are investigated. An error of the first kind is associated with a decoder that aims at decoding the entire coalition, whereas an error of the second kind is associated with a decoder whose target is to identify at least one member of the coalition. The setup of the game is similar to the one investigated in [32], with a few modifications.

- The first modification is that here, a more refined analysis of error exponents is performed while in [32], the capacity is the main quantity of interest (although some lower bounds on the error exponents are provided as well).

- The second difference between this paper and [32] concerns the distortion constraint imposed on the attacker. In this paper, the distortion between each of the fingerprinted copies observed by the attacker and the forgery he produces should be kept small, whereas in [32] it was sufficient to maintain a small amount of distortion between the forgery and one of the observed fingerprinted copies. This difference stems from the need to prevent an attack strategy that almost ignores one of its inputs, thereby rendering the goal of identifying the entire coalition impossible. Moreover, although ignoring one of the observed fingerprinted copies is a “legitimate” attack, it makes no sense that an effort will be made by the attacker to enlarge his coalition (by purchasing as many legal copies as possible), and then, at the end, ignore some of them.

We define the the achievable error exponent of the first/second kind as a number E such that there exists a random CC scheme, whose asymptotic performance in terms of average probability of error (in the logarithmic scale) of the first/second kind is given by E, assuming the attacker knows the encoding-decoding scheme and can adopt the worst-case strategy associated with the kind of error of interest. Single-letter lower and upper bounds are provided for the error exponents of the first and second kinds. In the case of an error of the first kind, the lower and upper bounds coincide, while when the error of the second kind is concerned, they coincide for a certain range of low rates. Another important result is that we show that one can use a universal decoder (used also in [32]) that is asymptotically optimal for the error of the first kind as well as an error of the second kind. We also deduce lower bounds on the capacities of the fingerprinting games corresponding to the two kinds of errors.

The paper is organized as follows: In Section 2, some notation conventions are defined. A statement of the problem, which is relevant to the entire paper, is given in Section 3. In Section 4 we describe our main results concerning the set of achievable error exponents, and Section 5 is devoted to a discussion of the results. Finally, the proofs of the theorems appear in Sections 6 and 7 and proofs of lemmas are deferred to Section 8.

2 Notation and Definitions

Henceforth, we adopt the following notation conventions. Random variables (RV's) will be denoted by capital letters, while their realizations will be denoted by the respective lower case letters. Random vectors of dimension n will be denoted by boldface letters. Thus, for example, if $\mathbf{X}$ denotes a random vector $(X_1, \ldots, X_n)$, then $\mathbf{x} = (x_1, \ldots, x_n)$ will designate a specific sample value of $\mathbf{X}$. The alphabet of a scalar RV, X, will be designated by the corresponding caligraphic letter $\mathcal{X}$. The n-fold Cartesian power of a generic alphabet $\mathcal{A}$, that is, the set of all n-vectors over $\mathcal{A}$, will be denoted $\mathcal{A}^n$.

The set of probability mass functions (pmf's), defined on a alphabet $\mathcal{X}$, will be denoted by $\mathcal{P}(\mathcal{X})$, and the set of conditional pmf's from $\mathcal{U}$ to $\mathcal{X}$ will be denoted $\mathcal{P}(\mathcal{X}|\mathcal{U})$, i.e.,

$$\mathcal{P}(\mathcal{X}|\mathcal{U}) = \left\{ P(X|U) : \; P(x|u) \geq 0, \; \sum_{x' \in \mathcal{X}} P(x'|u) = 1, \; \forall (u,x) \in \mathcal{U} \times \mathcal{X} \right\}. \tag{1}$$

The notation $\mathbf{1}\{A\}$, where A is an event, will designate the indicator function of A, i.e., $\mathbf{1}\{A\} = 1$ if A occurs, and $\mathbf{1}\{A\} = 0$ otherwise. We adopt the convention that if a set T is empty, then $\min_{t \in T} f(t) = \infty$, and similarly, $\max_{t \in T} f(t) = -\infty$. The notation $c_n \doteq d_n$, for two sequences $\{c_n\}_{n \geq 1}$ and $\{d_n\}_{n \geq 1}$, will express asymptotic equality in the exponential scale, i.e., $\lim_{n \to \infty} \frac{1}{n} \log \frac{c_n}{d_n} = 0$. Similarly, $c_n \stackrel{\cdot}{\geq} d_n$ will stand for $\liminf_{n \to \infty} \frac{1}{n} \log \frac{c_n}{d_n} \geq 0$, and so on.

The empirical pmf, induced by a vector $\mathbf{x} \in \mathcal{X}^n$, is the vector $\hat{P}_{\mathbf{x}} = \left\{ \hat{P}_{\mathbf{x}}(a), \; a \in \mathcal{X} \right\}$, where $\hat{P}_{\mathbf{x}}(a)$ is the relative frequency of the letter a in the vector $\mathbf{x}$. The set of empirical pmf's induced by all n-vectors in $\mathcal{X}^n$ will be denoted by $\mathbb{P}_n(\mathcal{X})$, i.e.,

$$\mathbb{P}_n(\mathcal{X}) = \left\{ \hat{P}_{\mathbf{x}} \right\}_{\mathbf{x} \in \mathcal{X}^n}. \tag{2}$$

The type class $T_{\mathbf{x}}$ (or $T(\hat{P}_{\mathbf{x}})$) is the set of n-vectors $\tilde{\mathbf{x}}$ such that $\hat{P}_{\tilde{\mathbf{x}}} = \hat{P}_{\mathbf{x}}$. Similarly, the joint empirical pmf induced by two n-vectors, $\mathbf{x}, \mathbf{y}$, is the vector $\hat{P}_{\mathbf{xy}} = \left\{ \hat{P}_{\mathbf{xy}}(a,b), \; a \in \mathcal{X}^n, b \in \mathcal{Y}^n \right\}$, where $\hat{P}_{\mathbf{xy}}(a,b)$ is the relative frequency of $(x_i, y_i) = (a,b)$. The type class $T_{\mathbf{xy}}$ (or $T(\hat{P}_{\mathbf{xy}})$) is the set of all pairs of n-vectors $\tilde{\mathbf{x}} \in \mathcal{X}^n, \tilde{\mathbf{y}} \in \mathcal{Y}^n$, such that $\hat{P}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}} = \hat{P}_{\mathbf{xy}}$. The conditional type class $T_{\mathbf{y}|\mathbf{x}}$ (or $T_{\mathbf{x}}(\hat{P}_{\mathbf{y}|\mathbf{x}})$), for a given $\mathbf{x}$, is the set of all n-vectors $\tilde{\mathbf{y}} \in \mathcal{Y}^n$ such that $T_{\mathbf{x}\tilde{\mathbf{y}}} = T_{\mathbf{xy}}$, and the conditional empirical pmf $\hat{P}_{\mathbf{y}|\mathbf{x}}$ is defined by $\hat{P}_{\mathbf{y}|\mathbf{x}}(y|x) = \frac{\hat{P}_{\mathbf{xy}}(x,y)}{\hat{P}_{\mathbf{x}}(x)}$, $\forall x \in \mathcal{X} : \hat{P}_{\mathbf{x}}(x) > 0$.

For a given empirical pmf $\hat{P} \in \mathbb{P}_n(\mathcal{A})$, define the set of conditional empirical pmf's,

$$\mathbb{P}_n(\mathcal{B}, \hat{P}) = \left\{ P \in \mathcal{P}(\mathcal{B}|\mathcal{A}) : \; n\hat{P}(a)P(b|a) \text{ is an integer for all } (a,b) \in \mathcal{A} \times \mathcal{B} \right\}. \tag{3}$$

The variational distance between two pmfs P and P', defined on the same set $\mathcal{A}$, will be denoted by $||P - P'||$, i.e.,

$$||P - P'|| = \sum_{a \in \mathcal{A}} |P(a) - P'(a)|. \tag{4}$$

Information-theoretic quantities, such as the entropy of the random variable X, whose pmf is P, will be denoted by either $H_P(X)$ or $H(P)$ interchangeably. Similarly, the mutual information between X and Y given U, with joint pmf P will be denoted by $I_P(X;Y|U)$, etc. The divergence, or Kullback-Leibler distance, between two pmf's P and Q on $\mathcal{A}$, where $|\mathcal{A}| < \infty$, is defined as $D(P\|Q) = \sum_{a\in\mathcal{A}} P(a)\log\frac{P(a)}{Q(a)}$, where we use the convention that $0\log 0 = 0$ and $p\log\frac{p}{0} = \infty$.

Information-theoretic quantities governed by empirical measures induced by the n-vectors $\mathbf{u}, \mathbf{x}, \mathbf{y}$, will have a special notation, e.g.,

$$\begin{aligned}\hat{H}_{\mathbf{x}} &\triangleq H_{\hat{P}_{\mathbf{x}}}(X)\\ \hat{I}_{\mathbf{x};\mathbf{y}|\mathbf{u}} &\triangleq I_{\hat{P}_{\mathbf{uxy}}}(X;Y|U).\end{aligned} \tag{5}$$

The notation $U \leftrightarrow X \leftrightarrow Y$ will signify that the RV's U, X, Y, in this order, form a Markov chain.

We shall have a particular interest in strongly exchangeable channels. A *strongly exchangeable channel* is defined as a conditional distribution $P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ with input alphabet $(\mathcal{X}^n)^2$ and output alphabet $\mathcal{Y}^n$ for which, for every $\mathbf{x}' \in \mathcal{X}^n, \mathbf{x} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^n$ and every permutation π of $\{1, \ldots, n\}$,

$$P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}\mathbf{x}') = P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\pi\mathbf{y}|\pi\mathbf{x}\pi\mathbf{x}'). \tag{6}$$

Obviously, every DMC is strongly exchangeable.

3 Statement of the Problem

As mentioned earlier, the setup of the game considered in this paper resembles the one investigated in [32] with several modifications discussed in the Introduction. Nevertheless, for the sake of completeness, we include the entire description of the modified game here.

Let $\mathcal{U}, \mathcal{X}$ and $\mathcal{Y}$ be finite sets designating the alphabets of a covertext symbol, fingerprinted symbol and forgery symbol, respectively. For convenience, we assume[3] $\mathcal{U} = \mathcal{X}$. Let $\mathbf{U}$ designate the random covertext sequence within which the fingerprints will be hidden. The n-vector $\mathbf{U}$ is composed of n i.i.d. RV's whose joint pmf is denoted by P_U^n with single-letter marginal pmf P_U, and we shall assume that $\min_{u\in\mathcal{U}} P_U(u) > 0$. The fingerprinter creates, at random, $M = \lfloor 2^{nR} \rfloor$ fingerprinted versions of the covertext, denoted $\mathbf{X}_i$, $i = 1, \ldots, M$, and will be referred to as a codebook.

A secret key, K_n, is an RV, independent of the fingerprints and the covertext, known to both the encoder and decoder, but unknown to the attacker. Let $\mathcal{K}_n$ stand for the alphabet of K_n, and let P_{K_n} stand for the distribution of K_n.

Definition 1. *A rate-R fingerprinting encoder of block-length n is a function which maps the secret key realization k_n, the covertext data $\mathbf{u}$, and the watermark message $m \in \mathcal{M}_n$ into*

[3] It is natural to make this assumption, because one of the things one wants to keep secret is the very existence of fingerprints. If $\mathcal{U} \neq \mathcal{X}$, it would be immediately apparent that the image or signal is fingerprinted.

a stegotext vector (or, fingerprinted vector) $\mathbf{x}$*, i.e.,*

$$f_n : \mathcal{K}_n \times \mathcal{U}^n \times \mathcal{M}_n \to \mathcal{X}^n, \tag{7}$$

where

$$\mathcal{M}_n = \{1, \ldots, M\}. \tag{8}$$

Having created the fingerprinted copies $\mathbf{X}_i$, $i \in \mathcal{M}_n$, the encoder distributes them arbitrarily to M different users. We assume that the distributed copies should meet the following distortion constraint w.r.t. a given distortion level D_1, i.e.,

$$\Pr\{d_1(\mathbf{U}, \mathbf{X}_i) \le nD_1\} = 1 \quad \forall i \in \mathcal{M}_n, \tag{9}$$

where $d_1 : \mathcal{U} \times \mathcal{X} \to \mathbb{R}_+$ denotes a single-letter distortion measure and $d_1(\mathbf{u}, \mathbf{x}) = \sum_{j=1}^n d_1(u_j, x_j)$ for $\mathbf{u} \in \mathcal{U}^n$ and $\mathbf{x} \in \mathcal{X}^n$.

Let W_a and W_b be the two[4] users that take part in the coalition. It is assumed that W_a, W_b are independent and both uniformly distributed over the set $\mathcal{M}_n$. Let $\mathbf{X}_a$ and $\mathbf{X}_b$ stand for the two corresponding sequences of fingerprinted data received by the users W_a and W_b. Namely, $\mathbf{X}_a$ and $\mathbf{X}_b$ are available to the attacker who creates the forgery, $\mathbf{Y}$, as the output of an attack channel $P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}$ from $\mathcal{X}^n \times \mathcal{X}^n$ to $\mathcal{Y}^n$.

Let $d_2 : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ denote another single-letter distortion measure, and denote $d_2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n d_2(x_i, y_i)$ for $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{y} \in \mathcal{Y}^n$. The attacker is assumed to meet the following distortion constraint[5]

$$\Pr\{d_2(\mathbf{X}_a, \mathbf{Y}) \le nD_2 \text{ and } d_2(\mathbf{X}_b, \mathbf{Y}) \le nD_2\} \;=\; 1. \tag{10}$$

Denote by $\mathcal{P}_n^{d_2}$ the set of channels satisfying (10).

The decoder observes the realization of the secret key (and thus, knows the particular codebook that has been drawn), the covertext $\mathbf{U}$ and the forgery $\mathbf{Y}$ and its aim is to detect members of the coalition. Thus, it is a function $\phi_n : \mathcal{K}_n \times \mathcal{U}^n \times \mathcal{Y}^n \to \mathcal{M}_n^2$.

As mentioned earlier, we analyze two kinds of errors: the first refers to a decoder that aims at decoding the two messages (detecting the entire coalition), and the second refers to a decoder that is less ambitious and is satisfied with correct decoding of at least one of the messages. We refer to the resulting errors as *error of the first kind* and *error of the second kind*, respectively. The quadruple $F_n = (P_{K_n}, f_n, \mathcal{M}_n, \phi_n)$ will be referred to as *a rate* R *randomized fingerprinting code.*

Definition 2. *Let* $\mathcal{N}_n(D_1)$ *be the set of mappings* $\eta : \mathcal{U}^n \to \mathcal{P}(\mathcal{X}|\mathcal{U})$*, s.t.*

$$E_{\hat{P}_{\mathbf{u}} \times \eta(\mathbf{u})} d_1(U, X) \le D_1 \quad \forall \mathbf{u} \in \mathcal{U}^n, \tag{11}$$

and $\eta(\mathbf{u}) \in \mathbb{P}_n(\mathcal{X}, \hat{P}_{\mathbf{u}})$*,* $\forall \mathbf{u} \in \mathcal{U}^n$*.*

[4] In the general fingerprinting game, the coalition may have more than two members. However, for the sake of simplicity, we focus on the case of two coalition members. The results can be extended to a general coalition size.

[5] In [32], the constraint was $\Pr\{d_2(\mathbf{X}_a, \mathbf{Y}) \le nD_2 \text{ and } d_2(\mathbf{X}_b, \mathbf{Y}) \le nD_2\} = 1$.

We shall focus on the following subclass of fingerprinting codes:

Definition 3. *A rate-R constant composition (CC) fingerprinting code of block-length n is a code with the following structure of a secret key and encoder: the fingerprinted copies $\mathbf{X}_i$, $i \in \mathcal{M}_n$, are drawn independently given $\mathbf{U}$, uniformly over a single conditional type class $T_{\mathbf{x}|\mathbf{u}}$ for all $\mathbf{u}$. The choice of the conditional type class is defined by a mapping*[6] $\eta \in \mathcal{N}_n(D_1)$, *i.e., it is given by* $T_{\mathbf{u}}(\eta(\mathbf{u}))$.

In the sequel, we shall use the abbreviation

$$T_\eta(\mathbf{u}) = T_{\mathbf{u}}(\eta(\mathbf{u})). \tag{12}$$

Denote by $\mathcal{F}_n^{d_1}(R)$ the set of rate-R randomized CC fingerprinting codes induced by a mapping $\eta_n \in \mathcal{N}_n(D_1)$. A code $F_n \in \mathcal{F}_n^{d_1}(R)$ is therefore defined by the triple (η_n, R, ϕ_n).

The fact that we focus on the wide class of CC fingerprinting encoders can be justified by practical considerations. As explained in [32], any practical randomized encoder should have some enumeration mechanism, where one first randomly selects a number under the uniform distribution in some range (in particular, an integer according to the key), and then this number is mapped to a codeword (given $\mathbf{U}$). It is desired then that to implement this mapping, one should not need (exponentially) large tables but can use a simple function. It is well known that there are indeed simple ways to enumerate sequences which belong to the same type class. See, for example, [10], where such an enumeration method is proposed, and the same idea can be easily extended to conditional type classes.

For a given realization of a secret key, K_n, let the output of the decoder be given by $\hat{W} = (\hat{W}_1, \hat{W}_2) = \phi_n(K_n, \mathbf{U}, \mathbf{Y})$. An error of the first kind occurs when not all coalition members are correctly detected[7] and an error of the second kind occurs if no coalition member is correctly detected by the decoder, hence, the average probability of error of the first and second kinds are given by

$$\begin{aligned}
&P_e^{(1)}\left(F_n, P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}\right) \triangleq \\
&\Pr\left\{\left\{\hat{W} \neq (W_a, W_b) \text{ and } \hat{W} \neq (W_b, W_a)\right\} \text{ or } \left\{W_a = W_b \text{ and } \hat{W}_1 \neq W_a \text{ and } \hat{W}_2 \neq W_a\right\}\right\} \\
&P_e^{(2)}\left(F_n, P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}\right) \triangleq \Pr\left\{\hat{W}_1 \neq W_a \text{ and } \hat{W}_1 \neq W_b \text{ and } \hat{W}_2 \neq W_a \text{ and } \hat{W}_2 \neq W_b\right\},
\end{aligned} \tag{13}$$

respectively, where the probability is induced by the covertext, the members of the coalition, the ensemble of all possible codebooks, and the action of the attack channel $P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}$, when the randomized code F_n is employed.

Definition 4. *An achievable rate R w.r.t. error of the first kind is one for which there exists a sequence $F_n \in \mathcal{F}_n^{d_1}(R)$, $n \geq 1$ such that*
$\limsup_{n\to\infty} \sup_{P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b} \in \mathcal{P}_n^{d_2}} P_e^{(1)}\left(F_n, P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}\right) = 0.$

[6] We require that $\eta \in \mathcal{N}_n(D_1)$ in order to consider conditional types such that constraint (9) is met.

[7] In the rare case where $W_a = W_b$, it is sufficient that either $\hat{W}_1 = W_a$ or $\hat{W}_2 = W_a$.

Definition 5. *The CC capacity of the private fingerprinting game w.r.t. a error of the first kind $C^{(1)}(P_U, D_1, D_2)$ is defined as the supremum of all achievable rates.*

Similar definitions apply for the achievable rate and the CC capacity w.r.t. error of the second kind $C^{(2)}(P_U, D_1, D_2)$. In fact, the capacity is a function of the distortion levels D_1, D_2 and the covertext symbol distribution P_U.

Define the negative normalized log error probability of the fingerprinting game w.r.t. error of the first and second kind when the code $F_n = (\eta_n, \phi_n, R)$ is applied

$$e_n^{(i)}(P_U, D_1, D_2, \eta_n, \phi_n, R) \triangleq -\frac{1}{n} \log P_e^{(i)}(\eta_n, \phi_n, R, P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}), \tag{14}$$

i=1,2, respectively.

Definition 6. *The error exponents of the fingerprinting game w.r.t. error of the first and second kind at rate R are defined by*

$$e^{(i)}(P_U, D_1, D_2, R) \triangleq \liminf_{n\to\infty} \max_{\eta_n \in \mathcal{N}_n(D_1), \phi_n} \min_{P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b} \in \mathcal{P}_n^{d_2}} e_n^{(i)}(P_U, D_1, D_2, \eta_n, \phi_n, R), \tag{15}$$

$i = 1, 2$, *respectively.*

Our main goal in this paper is to establish a closed form expression for the error exponents of the fingerprinting game w.r.t. error of the first and second kinds at rate R.

4 Main Results

For a given measure $P_{\tilde{U}X} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$, define the set

$$\mathcal{P}_{d_2}(P_{\tilde{U}X}, D_2) \triangleq \\ \left\{P_{\tilde{X}Y|\tilde{U}X} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}|\mathcal{U} \times \mathcal{X}) :\ P_{X|\tilde{U}} = P_{\tilde{X}|\tilde{U}}, \max\{Ed_2(X,Y), Ed_2(\tilde{X},Y)\} \le D_2\right\}, \tag{16}$$

where the conditional measures $P_{X|\tilde{U}}$ and $P_{\tilde{X}|\tilde{U}}$ are the appropriate marginals of $P_{\tilde{U}X} \times P_{\tilde{X}Y|\tilde{U}X}$, and the expectations are w.r.t. $P_{\tilde{U}X} \times P_{\tilde{X}Y|\tilde{U}X}$.

For a given pmf $P = P_{\tilde{U}X\tilde{X}Y} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}^2 \times \mathcal{Y})$ define the quantities

$$\begin{aligned} &\epsilon_a(P_{\tilde{U}X\tilde{X}Y}, R) \\ &\triangleq \left|\min\left\{I_P(X; \tilde{X}Y|\tilde{U}) - R\ ,\ I_P(\tilde{X}; XY|\tilde{U}) - R\ ,\ I_P(\tilde{X}; Y|\tilde{U}) + I_P(X; \tilde{X}Y|\tilde{U}) - 2R\right\}\right|^+ \\ &\epsilon_b(P_{\tilde{U}X\tilde{X}Y}, R) \triangleq \left|I_P(\tilde{X}; Y|\tilde{U}) + I_P(X; \tilde{X}Y|\tilde{U}) - 2R\right|^+ \end{aligned} \tag{17}$$

Define the following quantities for $i = \{a, b\}$

$$\begin{aligned} &E_i(P_U, D_1, D_2, R) \\ &\triangleq \min_{P_{\tilde{U}}} \max_{P_{X|\tilde{U}}} \min_{P_{\tilde{X}Y|\tilde{U}X}} \left[D(P_{\tilde{U}X\tilde{X}Y} || P_U \times P_{X|\tilde{U}} \times P_{\tilde{X}|\tilde{U}} \times P_{Y|X\tilde{X}}) + \epsilon_i(P_{\tilde{U}X\tilde{X}Y}, R)\right], \end{aligned} \tag{18}$$

where the outmost minimization is over $P_{\tilde{U}} \in \mathcal{P}(\mathcal{U})$, the maximization is over $P_{X|\tilde{U}}$ such that $Ed_1(\tilde{U},X) \leq D_1$, and the inner minimization is over $P_{\tilde{X}Y|\tilde{U}X} \in \mathcal{P}_{d_2}(P_{\tilde{U}X}, D_2)$. The measure $P_{Y|X\tilde{X}}$ is the appropriate marginal distribution induced by $P_{\tilde{U}X\tilde{X}Y}$. Note that

$$D(P_{\tilde{U}X\tilde{X}Y}||P_U \times P_{X|\tilde{U}} \times P_{\tilde{X}|\tilde{U}} \times P_{Y|X\tilde{X}}) = D(P_{\tilde{U}}||P_U) + I_P(X;\tilde{X}|\tilde{U}) + I_P(\tilde{U};Y|X\tilde{X}). \tag{19}$$

For convenience we denote

$$D(P_{\tilde{U}X\tilde{X}Y}, P_U) = D(P_{\tilde{U}X\tilde{X}Y}||P_U \times P_{X|\tilde{U}} \times P_{\tilde{X}|\tilde{U}} \times P_{Y|X\tilde{X}}). \tag{20}$$

The following theorem provides a single-letter expression of the error of the first kind.

Theorem 1. *For all P_U, D_1, D_2, R*

$$e^{(1)}(P_U, D_1, D_2, R) \quad = \quad E_a(P_U, D_1, D_2, R). \tag{21}$$

The proof of Theorem 1 appears in Section 6. It is composed of a lower bound and an upper bound on $e^{(1)}(P_U, D_1, D_2, R)$ which coincide.

As for the error of the second kind, define:

$$\begin{aligned}\tilde{E}_b(P_U, D_1, D_2, R) \quad &\triangleq \quad \min_{P_{\tilde{U}}} \max_{P_{X|\tilde{U}}} \min_{P_{\tilde{X}Y|\tilde{U}X}} \\ &\left[D(P_{\tilde{U}X\tilde{X}Y}||P_U \times P_{X|\tilde{U}} \times P_{\tilde{X}|\tilde{U}} \times P_{Y|X\tilde{X}}) + \epsilon_b(P_{\tilde{U}X\tilde{X}Y}, R)\right], \end{aligned} \tag{22}$$

where the outmost minimization is over $P_{\tilde{U}} \in \mathcal{P}(\mathcal{U})$, the maximization is over $P_{X|\tilde{U}}$ such that $Ed_1(\tilde{U},X) \leq D_1$, and the inner minimization is over

$$P_{\tilde{X}Y|\tilde{U}X} \in \mathcal{P}_{d_2}(P_{\tilde{U}X}, D_2) : \ \min\{I_P(X;\tilde{X}Y|\tilde{U}), I_P(\tilde{X};XY|\tilde{U})\} \geq R.$$

Let $R_0(P_U, D_1, D_2)$ be the lowest rate for which the constraint
$\min\{I_P(X;\tilde{X}Y|\tilde{U}), I_P(\tilde{X};XY|\tilde{U})\} \geq R$ appearing in the minimization becomes effective. The following theorem provides lower and upper bounds on the error exponent of the second kind.

Theorem 2. *For all P_U, D_1, D_2, R,*

$$E_b(P_U, D_1, D_2, R) \leq e^{(2)}(P_U, D_1, D_2, R) \leq \tilde{E}_b(P_U, D_1, D_2, R), \tag{23}$$

with equality whenever $R \in [0, R_0(P_U, D_1, D_2)]$.

The proof of Theorem 2 appears in Section 7.

The term $\epsilon_i(P_{\tilde{U}X\tilde{X}Y}, R)$ which appears in $E_i(P_U, D_1, D_2, R)$, can be interpreted as the error exponent conditioned on the event that $(\mathbf{U}, \mathbf{X}, \tilde{\mathbf{X}}, \mathbf{Y})$ lies within the type-class corresponding to $P_{\tilde{U}X\tilde{X}Y}$, and the term $D(P_{\tilde{U}X\tilde{X}Y}\|P_U \times P_{X|\tilde{U}} \times P_{\tilde{X}|\tilde{U}} \times P_{Y|X\tilde{X}})$ is a result of the averaging over the types.

5 Discussion

The asymptotically optimal decoder is the same decoder as in [32] denoted ϕ_n^* that is informed of $\mathbf{u}$ and the codebook, observes the forgery $\mathbf{y}$ and operates as follows:

$$(\hat{w}_1, \hat{w}_2) \;\; = \;\; \text{argmin}_{w_1,\, w_2 \neq w_1} |T_{\mathbf{x}_{w_1}\mathbf{x}_{w_2}|\mathbf{u}\mathbf{y}}|, \tag{24}$$

where ties are broken arbitrarily. Since this decoder achieves the lower bound on the average probability of error of the two kinds, it can be regarded as a universal decoder for the class of channels $\mathcal{P}_n^{d_2}$, w.r.t. random CC coding.

The fact that the maximizations over ϕ_n in (15) are performed separately for the two kinds of error exponents implies that the decoding rule can be chosen to minimize the kind of error of interest. In spite of this fact, it turns out that the same decoder is asymptotically optimal for the two kinds of errors, regardless of the encoding scheme.

Since $|T_{\mathbf{x}\mathbf{x}'|\mathbf{u}\mathbf{y}}| \doteq e^{n\hat{H}_{\mathbf{x}\mathbf{x}'|\mathbf{u}\mathbf{y}}}$, the alternative decoder which is also asymptotically optimal is given by

$$(\hat{w}_1, \hat{w}_2) \;\; = \;\; \text{argmin}_{w_1,\, w_2 \neq w_1} \hat{H}_{\mathbf{x}_{w_1}\mathbf{x}_{w_2}|\mathbf{u}\mathbf{y}}. \tag{25}$$

As the value of maxmin exponent is determined by the dominant joint type class (of the covertext sequence, the two fingerprinted copies of the coalition members and the forgery), an attack channel denoted $P^*_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ that assigns equal probability to all the conditional type classes $T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}$ such that the distortion constraint is not violated and is uniform within each conditional type class is introduced. Namely,

$$P^*_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}) = \frac{\mathbf{1}\left\{\max\{d_2(\mathbf{x},\mathbf{y}), d_2(\tilde{\mathbf{x}},\mathbf{y})\} \leq nD_2\right\}}{c_{n,\mathbf{x},\tilde{\mathbf{x}}}|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|} \tag{26}$$

where $c_{n,\mathbf{x},\tilde{\mathbf{x}}}$ is the appropriate polynomial normalization factor. This channel is shown to be a worst-case attack channel w.r.t. the error exponent of the first kind, and it is also used in the derivation of the upper bound on the error exponent of the second kind $\tilde{E}_b(P_U, D_1, D_2, R)$ (in (23)).

The proof of Theorems 1 and 2 involves the following lemma (proved independently also in [28]) whose proof appears in Section 8.

Lemma 1. *For all $a \in [0,1]$ and every integer $M \geq 1$,*

$$\frac{1}{2}\min\{1, Ma\} \leq 1 - (1-a)^M \leq \min\{1, Ma\}, \tag{27}$$

hence,

$$1 - (1-a)^M \doteq \min\{1, Ma\}. \tag{28}$$

This lemma implies that the union bound on the random coding error exponent is tight, and this lemma can also be used in other contexts such as in [13] to simplify the derivation.

The derivation performed in this paper provides us also with lower bounds on the capacities of the fingerprinting systems corresponding to the two kinds of errors, which are given by the smallest rates for which $E_i(P_U, D_1, D_2, R) = 0$, $i = a, b$. It is easily verified that this yields the following lower bounds

$$\begin{aligned} C^{(1)}(P_U, D_1, D_2) &\geq \max_{P_{X|U}} \min_{P_{Y|X\tilde{X}}} \min\left\{ I(\tilde{X}; Y|UX), I(X; Y|U\tilde{X}), \frac{1}{2} I(X\tilde{X}; Y|U) \right\} \\ C^{(2)}(P_U, D_1, D_2) &\geq \max_{P_{X|U}} \min_{P_{Y|X\tilde{X}}} \frac{1}{2} I(X\tilde{X}; Y|U), \end{aligned} \qquad (29)$$

where $U \sim P_U$, $\tilde{X}$ is an RV which is independent of X given U and satisfies $P_{\tilde{X}|U} = P_{X|U}$, and $U \leftrightarrow (X, \tilde{X}) \leftrightarrow Y$, the maximizations are over $\{P_{X|\tilde{U}} : E_{P_{\tilde{U}} P_{X|U}} d_1(U, X) \leq D_1\}$ and the minimizations are over $\left\{ P_{Y|\tilde{X}X} : \max\{E_{P_{X\tilde{X}} P_{Y|\tilde{X}X}} d_2(X, Y), E_{P_{X\tilde{X}} P_{Y|\tilde{X}X}} d_2(\tilde{X}, Y)\} \leq D_2 \right\}$.

The difference between the lower bound on the error exponent of the second kind, $E_b(P_U, D_1, D_2, R)$, and the lower bound on the error exponent of the MO decoder of [32] is that while in $E_b(P_U, D_1, D_2, R)$ (see (18)) the minimization is over $P_{\tilde{X}Y|\tilde{U}X} \in \mathcal{P}_{d_2}(P_{\tilde{U}X}, D_2)$, the bound in [32] includes a minimization over $P_{\tilde{X}Y|\tilde{U}X}$ such that $P_{X|\tilde{U}} = P_{X|U}$ and $\max\{Ed_2(X, Y), Ed_2(\tilde{X}, Y)\} \leq D_2$. This difference stems from the different distortion constraints imposed on the attacker.

In spite of the differences between the model of the ordinary MAC and the present scenario (see the discussion in the Introduction), the lower bound on the capacity $C^{(1)}(P_U, D_1, D_2)$ bears some resemblance to the capacity region of the MAC. The capacity region of the MAC given input distributions $P_1(X), P_2(\tilde{X})$ is given by the set of rate pairs (R_1, R_2), satisfying

$$0 \leq R_1 \leq I(X; Y|\tilde{X}) \; ; \; 0 \leq R_2 \leq I(\tilde{X}; Y|X) \; ; \; 0 \leq R_1 + R_2 \leq I(X, \tilde{X}; Y). \qquad (30)$$

In the case of two users who (use the same codebook and hence) have the same rate, $R_1 = R_2 = R$, the corresponding upper limit is $R \leq \min\{I(X; Y|\tilde{X}), I(\tilde{X}; Y|X), \frac{1}{2} I(X, \tilde{X}; Y)\}$, the minimizer depending on which of the above three lines is crossed first by the 45-degree line $R_1 = R_2$. The expression for the error exponent $E_a(P_U, D_1, D_2, R)$ also resembles the lower bound of the error exponent of the classical MAC (see [16]). That bound is given by

$$\begin{aligned} \min_{V_{UX\tilde{X}Y}} \Big[& D(V_{X\tilde{X}Y|U} \| P_{X|U} P_{\tilde{X}|U} P_{Y|X\tilde{X}} | P_U) \\ & + \max\{I(X; \tilde{X}Y|U) - R_1, I(\tilde{X}; XY|U) - R_2, I(\tilde{X}; Y|U) + I(X; \tilde{X}Y|U) - R_1 - R_2\} \Big], \end{aligned} \qquad (31)$$

where U is some auxiliary RV (used for time sharing) with distribution on some finite alphabet[8], $P_{X|U}$ and $P_{\tilde{X}|U}$ are conditional distributions (that can be optimized) defined on $\mathcal{P}(\mathcal{X}|\mathcal{U})$ and $\mathcal{P}(\tilde{\mathcal{X}}|\mathcal{U})$, respectively, with $\mathcal{X}, \tilde{\mathcal{X}}$ being the inputs alphabets of the channel, and the minimization is over $V_{UX\tilde{X}Y} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \tilde{\mathcal{X}} \times \mathcal{Y})$ with marginals $V_{UX} = P_{UX}$

[8] It is proved in [16] that the size of the alphabet of U can be 4 without loss of generality.

and $V_{U\tilde{X}} = P_{U\tilde{X}}$. The main differences between (31) and $E_a(P_U, D_1, D_2, R)$ stem from the distortion constraints imposed on both parties of the fingerprinting game and from the fact that we consider a coalition of two users sharing the same codebook and thus operating at the same rate. It should be noted that while U in $E_a(P_U, D_1, D_2, R)$ represents a covertext symbol, in [16] it stands for a time-sharing symbol. A modification of the derivation performed in this paper (the lower bound on $P_e^{(1)}\left(F_n, P_{\mathbf{Y}|\mathbf{X}_a\mathbf{X}_b}\right)$), can be used to show that the bound of [16] is tight in the random coding sense.

6 Proof of Theorem 1

6.1 Proof of the Direct Part of Theorem 1

The performance of the decoder ϕ_n^* (see (24)) will serve as an upper bound on the average probability of error attainable by the optimal decoder. The encoder is a CC fingerprinting code defined by a mapping $\eta \in \mathcal{N}_n(D_1)$ that satisfies the following condition

$$\eta(\mathbf{u}) = \eta(\pi\mathbf{u}) \tag{32}$$

for all $\mathbf{u} \in \mathcal{U}^n$ and every permutation π of $\{1, ..., n\}$. In other words, the channel from $\mathbf{U}$ to $\mathbf{X}$ is strongly exchangeable.

Without loss of generality[9] one can assume that the transmitted messages indices are $(1, 2)$. With a little abuse of notation, we shall denote by $\mathbf{P}_{\mathbf{u}}^M$ the joint pmf of $\{\mathbf{X}_i\}_{i=1}^M$ conditioned on the event $\mathbf{U} = \mathbf{u}$, and by $\mathbf{P}_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}^M$ the joint pmf of $\{\mathbf{X}_i\}_{i=1}^M$ conditioned on the event $\mathbf{U} = \mathbf{u}, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2$.

Assuming the covertext is $\mathbf{u}$, the codewords observed by the attacker are $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{y}$ is the forgery, the probability that the proposed decoder (24) fails to decode the entire coalition is given by

$$\begin{aligned}
& \Pr\left\{\text{error } |\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\right\} \\
= \; & \mathbf{P}_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}^M \left\{\exists (i,j) \neq (1,2) : \; |T_{\mathbf{x}_i,\mathbf{x}_j|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|\right\} \\
\doteq \; & \sum_{T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_1\mathbf{y}} : \, |T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_1\mathbf{y}}| \leq |T_{\mathbf{x}_2|\mathbf{u}\mathbf{x}_1\mathbf{y}}|} \mathbf{P}_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}^M \left\{\exists k \geq 3 : \; \mathbf{X}_k \in T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_1\mathbf{y}}\right\} \\
+ \; & \sum_{T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_2\mathbf{y}} : \, |T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_2\mathbf{y}}| \leq |T_{\mathbf{x}_1|\mathbf{u}\mathbf{x}_2\mathbf{y}}|} \mathbf{P}_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}^M \left\{\exists k \geq 3 : \; \mathbf{X}_k \in T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_2\mathbf{y}}\right\} \\
+ \; & \sum_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}} : \, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} \mathbf{P}_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}^M \left\{\exists i \geq 3, j \geq 3 : \; (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}\right\} \qquad (33)
\end{aligned}$$

$$(34)$$

[9] The case $w_1 = w_2$ can be treated similarly.

$$\begin{aligned}
&\doteq \max_{T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_1\mathbf{y}}:\, |T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_1\mathbf{y}}| \le |T_{\mathbf{x}_2|\mathbf{u}\mathbf{x}_1\mathbf{y}}|} \left[1 - \left(1 - \frac{|T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_1\mathbf{y}}|}{|T_\eta(\mathbf{u})|}\right)^{M-2}\right] \\
&+ \max_{T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_2\mathbf{y}}:\, |T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_2\mathbf{y}}| \le |T_{\mathbf{x}_1|\mathbf{u}\mathbf{x}_2\mathbf{y}}|} \left[1 - \left(1 - \frac{|T_{\mathbf{x}'|\mathbf{u}\mathbf{x}_2\mathbf{y}}|}{|T_\eta(\mathbf{u})|}\right)^{M-2}\right] \\
&+ \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \le |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2} \left\{\exists i \ge 3, j \ge 3 :\ (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}\right\} \\
&= \left[1 - \left(1 - \frac{|T_{\mathbf{x}_2|\mathbf{u}\mathbf{x}_1\mathbf{y}}|}{|T_\eta(\mathbf{u})|}\right)^{M-2}\right] + \left[1 - \left(1 - \frac{|T_{\mathbf{x}_1|\mathbf{u}\mathbf{x}_2\mathbf{y}}|}{|T_\eta(\mathbf{u})|}\right)^{M-2}\right] \\
&+ \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \le |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2} \left\{\exists i \ge 3, j \ge 3 :\ (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}\right\}, \qquad (35)
\end{aligned}$$

where the three summands in (33) are corresponding to (i) error only in $\hat{W}_2$, (ii) error only in $\hat{W}_1$, and (iii) error in both $\hat{W}_1$ and $\hat{W}_2$.

Next, we present a key lemma that will be used to evaluate (35) and the lower bound as well. Recall the abbreviation (12).

Lemma 2. *Let $\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y} \in \mathcal{U}^n \times \mathcal{X}^n \times \mathcal{X}^n \times \mathcal{Y}^n$ be given n-vectors such that $\mathbf{x} \in T_\eta(\mathbf{u}), \tilde{\mathbf{x}} \in T_\eta(\mathbf{u})$. The quantity*

$$P_M \triangleq P_M(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{u}, \mathbf{y}) = \mathbf{P}^M_{\mathbf{u}} \left\{\exists (i,j) \in \{1, ..., M\}^2 \text{ s.t. } (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}\right\} \qquad (36)$$

satisfies

$$\begin{aligned}
P_M &\le \min\{1, C_M\} && (37)\\
P_M &\ge q_{M-2} \cdot \frac{C_M}{1 + C_M} && (38)
\end{aligned}$$

where

$$\begin{aligned}
q_M &\triangleq q_M(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{u}, \mathbf{y}) = \left[1 - \frac{|T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}} \cup T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}}|}{|T_\eta(\mathbf{u})|}\right]^M, && (39)\\
C_M &\triangleq C_M(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{u}, \mathbf{y}) = \begin{pmatrix} M \\ 2 \end{pmatrix} \cdot \frac{|T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}|}{|T_\eta(\mathbf{u})|^2}. && (40)
\end{aligned}$$

The lemma is proved in Section 8.

Obviously, by (35) and (37) we have the upper bound

$$\begin{aligned}
&\mathbf{P}^M_{\mathbf{u}} \left\{\exists (i,j) \ne (1,2) :\ |T_{\mathbf{x}_i,\mathbf{x}_j|\mathbf{u}\mathbf{y}}| \le |T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}|\right\} \\
&\le 2(1 - q_{M-2}) + \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \le |T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}|} \mathbf{P}^M_{\mathbf{u}} \left\{\exists i \ge 3, j \ge 3 :\ (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}\right\} \\
&\doteq 1 - q_{M-2} + \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \le |T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}|} P_{M-2}(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y}) \\
&\le 1 - q_{M-2} + \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \le |T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}|} \min\{1, C_{M-2}(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y})\} \\
&= 1 - q_{M-2} + \min\{1, C_{M-2}\} \qquad (41)\\
&\doteq \max\{1 - q_{M-2}, \min\{1, C_{M-2}\}\}. \qquad (42)
\end{aligned}$$

where (41) follows since $C_M(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y})$ is increasing with $|T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}|$. Thus,

$$\begin{aligned} & \Pr\{\text{error of the proposed decoder } |\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\} \\ \overset{\cdot}{\leq} \;& \max\{1 - q_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}), \min\{1, C_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\}\}, \end{aligned} \tag{43}$$

Next, denote

$$a = a(\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}) \triangleq \frac{|T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}} \cup T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}\mathbf{y}}|}{|T_{\mathbf{x}|\mathbf{u}}|}, \tag{44}$$

and note that since $T_\eta(\mathbf{u}) = T_{\mathbf{x}|\mathbf{u}}$ and $q_M = (1-a)^M$, Lemma 1 provides the asymptotic behavior of $1 - q_M$ and implies,

$$1 - q_M \doteq \min\{1, Ma\}, \tag{45}$$

and consequently,

$$\begin{aligned} & \max\{1 - q_{M-2}\ ,\ \min\{1, C_{M-2}\}\} \\ \doteq\; & \max\{\min\{1, Ma\}\ ,\ \min\{1, C_M\}\} \\ =\; & \min\{1, \max\{Ma\ ,\ C_M\}\}. \end{aligned} \tag{46}$$

For the sake of convenience, we shall denote

$$\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) \triangleq \min\{1, \max\{Ma\ ,\ C_M\}\}. \tag{47}$$

We thus have established the upper bound

$$\begin{aligned} & \min_{\eta \in \mathcal{N}_n(D_1), \phi_n} \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}} \in \mathcal{P}_n^{d_2}} P_e^{(1)}(\eta, \phi_n, R, P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}) \\ \overset{\cdot}{\leq}\; & \min_{\eta \in \mathcal{N}_n(D_1), \phi_n} \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}} \in \mathcal{P}_n^{d_2}} \sum_{\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}} \Pr(\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}) \alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}), \end{aligned} \tag{48}$$

where

$$\Pr(\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}) = P_U^n(\mathbf{u}) \frac{\mathbf{1}\{\mathbf{x} \in T_\eta(\mathbf{u}), \tilde{\mathbf{x}} \in T_\eta(\mathbf{u})\}}{|T_\eta(\mathbf{u})|^2} P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}). \tag{49}$$

The next two lemmas, whose proofs appear in Section 8, conclude the proof of the direct part of Theorem 1. Denote by $\mathcal{N}_n^{ex}(D_1)$ the set of mappings $\eta \in \mathcal{N}_n(D_1)$ that satisfy (32).

Lemma 3. *For every $\eta \in \mathcal{N}_n^{ex}(D_1)$,*

$$\begin{aligned} & \min_{\phi_n} \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}} \in \mathcal{P}_n^{d_2}} P_e^{(1)}(\eta, \phi_n, R, P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}) \\ \overset{\cdot}{\leq}\; & \sum_{\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}} Pr(\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}) \sum_{\mathbf{y}:\ \max\{d_2(\mathbf{x}, \mathbf{y}), d_2(\tilde{\mathbf{x}}, \mathbf{y})\} \leq nD_2} \frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}}|}, \end{aligned} \tag{50}$$

where $Pr(\mathbf{u}, \mathbf{x}, \tilde{\mathbf{x}}) = P_U^n(\mathbf{u}) \frac{\mathbf{1}\{\mathbf{x} \in T_\eta(\mathbf{u}), \tilde{\mathbf{x}} \in T_\eta(\mathbf{u})\}}{|T_\eta(\mathbf{u})|^2}$.

Lemma 4.

$$
\begin{aligned}
&\min_{\eta\in\mathcal{N}_n^{ex}(D_1)}\sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}} Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}) \sum_{\mathbf{y}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\leq nD_2} \frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}}|}\\
=\;&\min_{\eta\in\mathcal{N}_n(D_1)}\sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}} Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}) \sum_{\mathbf{y}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\leq nD_2} \frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}}|} \qquad (51)\\
\doteq\;&\exp\left(-n\min_{P_{\tilde{U}}}\max_{P_{X|\tilde{U}}}\min_{P_{\tilde{X}Y|\tilde{U},X}}\left[D(P_{\tilde{U}X\tilde{X}Y},P_U)+\epsilon_a(P_{\tilde{U}X\tilde{X}Y},R)\right]\right), \qquad (52)
\end{aligned}
$$

where the outmost minimization is over $P_{\tilde{U}}\in\mathbb{P}_n(\mathcal{U})$, the maximization is over $P_{X|\tilde{U}}\in\mathbb{P}_n(\mathcal{X},P_{\tilde{U}}):\ Ed_1(\tilde{U},X)\leq D_1$, and the innermost minimization is over $P_{\tilde{X}Y|\tilde{U}X}\in\mathbb{P}_n(\mathcal{X}\times\mathcal{Y},P_{\tilde{U}X}):$ s.t. $P_{\tilde{X}|\tilde{U}}=P_{X|\tilde{U}}$ and $\max\{Ed_2(X,Y),Ed_2(\tilde{X},Y)\}\leq D_2$.

The gap between the r.h.s. of (52) and $E_a(P_U,D_1,D_2,R)$ is only in that in (52), the optimizations are over empirical measures while in (93) the optimizations are over sets of continuous measures. Due to the continuity considerations, as n tends infinity, the r.h.s. of (52) converges to $E_a(P_U,D_2,R)$.

6.2 Proof of the Converse Part of Theorem 1

When deriving a lower bound on the average probability of error, one can assume (a) that the attacker is constrained to use a strongly exchangeable channel and (b) that the channel is known at the decoder's side which can implement the maximum likelihood decoding rule. The next lemma will be used to establish a lower bound on the probability of error of the ML decoder under these assumptions.

Lemma 5. *For any fixed codebook and a known channel that is strongly exchangeable, the ML decoder assigns the same likelihood to two pairs of codewords that lie in the same conditional type given $(\mathbf{u},\mathbf{y})$.*

Proof. Let $\mathbf{B}_{\mathbf{u}}$ denote the codebook corresponding to $\mathbf{u}$, i.e., the collection of codewords $\{\mathbf{x}_i(\mathbf{u})\}_{i=1}^M$. Let m,m' be two message indices. Since $(\mathbf{u},\mathbf{B}_{\mathbf{u}},\mathbf{y})$ are known at the decoder, the ML decoder should maximize the quantity $\Pr(m,m'|\mathbf{u},\mathbf{B}_{\mathbf{u}},\mathbf{y})$ over all $m\in\mathcal{M}_n$ $m'\in\mathcal{M}_n$ $m\neq m'$. We have

$$
\begin{aligned}
&\Pr(m,m'|\mathbf{u},\mathbf{B}_{\mathbf{u}},\mathbf{y})\\
\overset{(a)}{=}\;&\frac{\Pr(m,m',\mathbf{u},\mathbf{B}_{\mathbf{u}})}{\Pr(\mathbf{u},\mathbf{B}_{\mathbf{u}},\mathbf{y})}\Pr(\mathbf{y}|m,m',\mathbf{u},\mathbf{B}_{\mathbf{u}})\\
\overset{(b)}{=}\;&\frac{\frac{1}{M^2}\Pr(\mathbf{u},\mathbf{B}_{\mathbf{u}})}{\Pr(\mathbf{u},\mathbf{B}_{\mathbf{u}},\mathbf{y})}P_{\mathbf{Y}|\mathbf{X},\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}_m(\mathbf{u}),\mathbf{x}_{m'}(\mathbf{u})) \qquad (53)
\end{aligned}
$$

where (a) follows from Bayes rule, (b) follows from Bayes rule and the fact that $\Pr(\mathbf{y}|\mathbf{u},m,m',\mathbf{B}_{\mathbf{u}})=\Pr(\mathbf{y}|\mathbf{x}_m(\mathbf{u}),\mathbf{x}_{m'}(\mathbf{u}))$, $\Pr(m,m')=\frac{1}{M^2}$, and $(\mathbf{u},\mathbf{B}_{\mathbf{u}})$ is independent of the message indices.

Hence, the ML decoder should maximize $P_{\mathbf{Y}|\mathbf{X},\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}_m(\mathbf{u}),\mathbf{x}_{m'}(\mathbf{u}))$ over all message indices $m \neq m'$. Since $P_{\mathbf{Y}|\mathbf{X},\mathbf{X}'}$ is strongly exchangeable, by definition (see (6)), if $(\mathbf{x}_{\tilde{m}}(\mathbf{u}),\mathbf{x}_{m''}(\mathbf{u})) \in T_{\mathbf{x}_m(\mathbf{u}),\mathbf{x}_{m'}(\mathbf{u})|\mathbf{y}}$ one has

$$P_{\mathbf{Y}|\mathbf{X},\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}_m(\mathbf{u}),\mathbf{x}_{m'}(\mathbf{u})) = P_{\mathbf{Y}|\mathbf{X},\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}_{\tilde{m}}(\mathbf{u}),\mathbf{x}_{m''}(\mathbf{u})) \tag{54}$$

and the lemma follows. □

Thus, the probability that the ML decoder fails (given $\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}$) is lower bounded as follows:

$$\begin{aligned} &\Pr\{\text{error of ML decoder, known exchangeable channel } |\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\ \} \\ \geq\ & \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists(i,j)\neq(1,2):\ (\mathbf{x}_i,\mathbf{x}_j)\in T_{\mathbf{x}_1\mathbf{x}_2|\mathbf{u}\mathbf{y}}\right\}. \end{aligned} \tag{55}$$

We can now use (38) to lower bound the probability of the event of interest.

$$\begin{aligned} &\mathbf{P}^M_{\mathbf{u}}\left\{\exists(i,j)\neq(1,2):\ (\mathbf{x}_i,\mathbf{x}_j)\in T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}\right\} \\ &= \mathbf{P}^M_{\mathbf{u}}\left\{\exists k\geq 3:\ \mathbf{X}_k\in\{T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}}\cup T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}\mathbf{y}}\} \text{ or } \exists(i,j), i,j\geq 3:\ (\mathbf{X}_i,\mathbf{X}_j)\in T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}\right\} \\ &\geq \max\left\{\mathbf{P}^M_{\mathbf{u}}\left\{\exists k\geq 3:\ \mathbf{X}_k\in\{T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}}\cup T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}\mathbf{y}}\}\right\}, \mathbf{P}^M_{\mathbf{u}}\left\{\exists(i,j), i,j\geq 3:\ (\mathbf{X}_i,\mathbf{X}_j)\in T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}\right\}\right\} \\ &= \max\{1-q_{M-2}, P_{M-2}\} \\ &\geq \max\left\{1-q_{M-2}, q_{M-4}\cdot\frac{C_{M-2}}{1+C_{M-2}}\right\} \\ &\geq \max\left\{1-q_{M-2}\ ,\ q_{M-2}\cdot\frac{C_{M-2}}{1+C_{M-2}}\right\}, \end{aligned} \tag{56}$$

where (56) is due to the fact that q_M is decreasing with M. Next, denote for convenience $q \triangleq q_{M-2}$ and $C = C_{M-2}$

$$\begin{aligned} & \max\left\{1-q\ ,\ q\cdot\frac{C}{1+C}\right\} \\ =\ & (1-q)\cdot\mathbf{1}\left\{1-q\geq q\cdot\frac{C}{1+C}\right\} \\ & +q\cdot\frac{C}{1+C}\cdot\mathbf{1}\left\{1-q<q\cdot\frac{C}{1+C}\right\} \\ =\ & (1-q)\cdot\mathbf{1}\left\{q\leq\frac{1+C}{1+2C}\right\} \\ & +q\cdot\frac{C}{1+C}\cdot\mathbf{1}\left\{q>\frac{1+C}{1+2C}\right\} \\ \geq\ & \left(1-\frac{1+C}{1+2C}\right)\cdot\mathbf{1}\left\{q\leq\frac{1+C}{1+2C}\right\} \\ & +\frac{1+C}{1+2C}\cdot\frac{C}{1+C}\cdot\mathbf{1}\left\{q>\frac{1+C}{1+2C}\right\} \\ =\ & \frac{C}{1+2C} \end{aligned} \tag{57}$$

and obviously,

$$\max\left\{1-q\ ,\ q\cdot\frac{C}{1+C}\right\}\geq(1-q),\tag{58}$$

thus, (57) and (58) imply

$$\begin{aligned}
& \max\left\{1-q_{M-2}\ ,\ q_{M-2}\cdot\frac{C_{M-2}}{1+C_{M-2}}\right\} \\
\geq\ & \max\left\{1-q_{M-2}\ ,\ \frac{C_{M-2}}{1+2C_{M-2}}\right\} \\
\geq\ & \max\left\{1-q_{M-2}\ ,\ \frac{1}{3}\cdot\min\left\{1,C_{M-2}\right\}\right\} \qquad (59)\\
\doteq\ & \max\left\{1-q_{M-2}\ ,\ \min\left\{1,C_{M-2}\right\}\right\}, \qquad (60)
\end{aligned}$$

where (59) follows because either $C_{M-2}\leq 1$ and then $\frac{C_{M-2}}{1+2C_{M-2}}\geq\frac{C_{M-2}}{3}$ or $C_{M-2}\geq 1$, and consequently $\frac{C_{M-2}}{1+2C_{M-2}}\geq\frac{1}{3}$.

The fact that the bounds (43) and (60) coincide yields that the average probability of error (given $\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}$) of the proposed decoder (24) achieves the lower bound on the average probability of error attainable by the ML decoder which corresponds to every strongly exchangeable channel that is known at the decoder.

We have thus established the lower bound

$$\begin{aligned}
& \min_{\eta\in\mathcal{N}_n(D_1),\phi_n}\ \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\in\mathcal{P}_n^{d_2}} P_e^{(1)}(\eta,\phi_n,R,P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}) \\
\overset{\cdot}{\geq}\ & \min_{\eta\in\mathcal{N}_n(D_1)}\ \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\in\mathcal{P}_n^{d_2,ex}}\ \sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}},\mathbf{y}}\Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}},\mathbf{y})\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}), \qquad (61)
\end{aligned}$$

where $\Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}},\mathbf{y})$ is as in (49) and

$$\mathcal{P}_n^{d_2,ex}\ =\ \left\{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\in\mathcal{P}_n^{d_2}\ :\ P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\text{ is strongly exchangeable}\right\}.\tag{62}$$

The gap between (48) and (61) stems only from the fact that the set over which the maximization is performed in the lower bound is $\mathcal{P}_n^{d_2,ex}$ while in the lower bound the set is $\mathcal{P}_n^{d_2}$. To bridge this gap, we introduce the attack channel $P^*_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ given in (26), and to conclude the proof we note that

$$\begin{aligned}
& \min_{\eta\in\mathcal{N}_n(D_1)}\ \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\in\mathcal{P}_n^{d_2,ex}}\ \sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}},\mathbf{y}}\Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}},\mathbf{y})\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) \\
\geq\ & \sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}},\mathbf{y}}\Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}})P^*_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}})\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) \\
\doteq\ & \sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}}\Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}})\sum_{\mathbf{y}:\ \max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\leq nD_2}\frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}}|}, \qquad (63)
\end{aligned}$$

applying Lemma 4 and its proceeding remark, the converse part of Theorem 1 follows.

7 Proof of Theorem 2

7.1 Proof of the Lower Bound of Theorem 2

When the error of the second kind is concerned, the probability that the proposed decoder (24) fails to decode at least one member of the coalition is given by

$$\begin{aligned}
& \Pr\left\{\text{error of the proposed decoder } |\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\right\} \\
= \; & \mathbf{P}^M_{\mathbf{u}\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3: \; |T_{\mathbf{x}_i,\mathbf{x}_j|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|\right\} \\
= \; & \sum_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3: \; (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}\right\} \\
\doteq \; & \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3: \; (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}\right\} \\
= \; & \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} P_{M-2}(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y}). \qquad (64)
\end{aligned}$$

We have,

$$\begin{aligned}
& \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} P_{M-2}(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y}) \\
\leq \; & \max_{T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}:\, |T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}| \leq |T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}|} \min\left\{1, C_{M-2}(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y})\right\} \\
= \; & \min\left\{1, C_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\right\} \qquad (65)
\end{aligned}$$

where the last step follows since $C_M(\mathbf{u}\mathbf{x}'\mathbf{x}''\mathbf{y})$ is increasing with $|T_{\mathbf{x}'\mathbf{x}''|\mathbf{u}\mathbf{y}}|$. Now we use a similar derivation to the one performed in Lemmas 3 and 4 (replacing $\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})$ defined in (47) by $\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) = \min\left\{1, C_M(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\right\}$), and this yields

$$\min_{\eta \in \mathcal{N}_n(D_1), \phi_n} \; \max_{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}} \in \mathcal{P}_n^{d_2}} P_e^{(1)}(\eta, \phi_n, R, P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}) \qquad (66)$$

$$\begin{aligned}
& \max_{\phi_n} e_n^{(2)}(P_U, D_2, \eta, \phi_n, R) \\
\dot{\geq} \; & \min_{\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}} \in \mathcal{P}_{d_2}^{(n)}(\eta, D_2)} \left[D(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}} || P_U \times \hat{P}_{\mathbf{x}|\mathbf{u}} \times \hat{P}_{\tilde{\mathbf{x}}|\mathbf{u}} \times \hat{P}_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}) + \epsilon_b(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}, R) \right]. \qquad (67)
\end{aligned}$$

Since $\eta \in \mathcal{N}_n(D_1)$, the mapping η is continuous, and thus, the r.h.s. of the above equality converges to $E_b(P_U, D_2, \eta, R)$. This concludes the proof of the l.h.s. of (23).

7.2 Proof of the Upper Bound of Theorem 2

Similarly to the argumentation used in the analysis of the error of the first kind, when deriving a lower bound on the probability of error of the second kind, one can assume (a) that the attacker is constrained to use strongly exchangeable channels and (b) that the channel is known at the decoder's side which can implement an optimal decoding rule w.r.t. the error of second kind. We next characterize this optimal decoder. Recall that Let $\mathbf{B_u}$ denotes the codebook corresponding to $\mathbf{u}$. First, note that given $(\mathbf{u}, \mathbf{y})$ and when the optimal decoder for the error of second kind is used, if there exists another pair $(\mathbf{x}_i, \mathbf{x}_j)$ where $i \geq 3$ and $j \geq 3$ such that $(\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u},\mathbf{y}}$ an error occur with high probability, because, the optimal decoder for the error of second kind minimizes

$$\Pr(m_1 \text{ not sent }, m_2 \text{ not sent}|\mathbf{u}, \mathbf{B_u}, \mathbf{y}) = 1 - \Pr(m_1 \text{ sent or } m_2 \text{ sent}|\mathbf{u}, \mathbf{B_u}, \mathbf{y}) \tag{68}$$

among all $m_1, m_2 \in \mathcal{M}_n$, $m_1 \neq m_2$, or alternatively, maximizes

$$\Pr(m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u}, \mathbf{y}) + \Pr(m_2 \text{ sent}|\mathbf{u}, \mathbf{B_u}, \mathbf{y}) - \Pr(m_1 \text{ sent}, m_2 \text{ sent }|\mathbf{u}, \mathbf{B_u}, \mathbf{y}) \tag{69}$$

multiplying by $\frac{\Pr(\mathbf{y}|\mathbf{u},\mathbf{B_u})}{\Pr(m_1 \text{ sent }|\mathbf{u},\mathbf{B_u})}$ we obtain

$$\begin{aligned}
&\frac{\Pr(\mathbf{y}, m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u})}{\Pr(m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u})} + \frac{\Pr(\mathbf{y}, m_2 \text{ sent}|\mathbf{u}, \mathbf{B_u})}{\Pr(m_1 \text{ sent}|\mathbf{u}, \mathbf{B_u})} - \frac{\Pr(\mathbf{y}, m_1 \text{ sent}, m_2 \text{ sent }|\mathbf{u}, \mathbf{B_u})}{\Pr(m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u})} \\
&= \frac{\Pr(\mathbf{y}, m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u})}{\Pr(m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u})} + \frac{\Pr(\mathbf{y}, m_2 \text{ sent}|\mathbf{u}, \mathbf{B_u})}{\Pr(m_2 \text{ sent}|\mathbf{u}, \mathbf{B_u})} - \frac{\frac{2}{M^2}}{(\frac{2}{M} - \frac{1}{M^2})} \cdot \frac{\Pr(\mathbf{y}, m_1 \text{ sent}, m_2 \text{ sent }|\mathbf{u}, \mathbf{B_u})}{\Pr(m_1 \text{ sent}, m_2 \text{ sent }|\mathbf{u}, \mathbf{B_u})},
\end{aligned} \tag{70}$$

where the last step follows since $\Pr(m_1 \text{ sent }|\mathbf{u}, \mathbf{B_u}) = \Pr(m_2 \text{ sent }|\mathbf{u}, \mathbf{B_u}) = \frac{2}{M} - \frac{1}{M^2}$ and $\Pr(m_1 \text{ sent}, m_2 \text{ sent }|\mathbf{u}, \mathbf{B_u}) = \frac{2}{M^2}$. Hence, the optimal decoder maximizes

$$\begin{aligned}
&\Pr(\mathbf{y}|\mathbf{u}, \mathbf{B_u}, m_1 \text{ sent }) + \Pr(\mathbf{y}|\mathbf{u}, \mathbf{B_u}, m_2 \text{ sent}) - \frac{2}{2M-1} \cdot \Pr(\mathbf{y}|\mathbf{u}, \mathbf{B_u}, m_1 \text{ sent }, m_2 \text{ sent }) \\
&= \Pr(\mathbf{y}|\mathbf{u}, \mathbf{B_u}, \mathbf{x}_{m_1}) + \Pr(\mathbf{y}|\mathbf{u}, \mathbf{B_u}, \mathbf{x}_{m_2}) - \frac{2}{2M-1} \cdot \Pr(\mathbf{y}|\mathbf{x}_{m_1}, \mathbf{x}_{m_2}) \\
&= \sum_{m'} \frac{1}{M} \left[P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}_{m_1}, \mathbf{x}_{m'}) + P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}_{m'}, \mathbf{x}_{m_1},)\right] \\
&+ \sum_{m'} \frac{1}{M} \left[P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}_{m_2}, \mathbf{x}_{m'}) + P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}_{m'}, \mathbf{x}_{m_2})\right] \\
&- \frac{1}{2M-1} \cdot \left[P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}_{m_1}, \mathbf{x}_{m_2}) + P_{\mathbf{Y}|\mathbf{X}\mathbf{X}'}(\mathbf{y}|\mathbf{x}_{m_2}, \mathbf{x}_{m_1})\right] \\
&\triangleq f_{\mathbf{u},\mathbf{B_u},\mathbf{y}}(\mathbf{x}_{m_1}, \mathbf{x}_{m_2}).
\end{aligned} \tag{71}$$

Therefore, even for the optimal decoder (w.r.t. average probability of error of the second kind) one has

$$\begin{aligned}
&\Pr\{\text{error }|\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\} \\
&= \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3 : \; f_{\mathbf{u},\mathbf{B_u},\mathbf{y}}(\mathbf{x}_i, \mathbf{x}_j) \geq f_{\mathbf{u},\mathbf{B_u},\mathbf{y}}(\mathbf{x}_1, \mathbf{x}_2)|\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\right\} \\
&\geq \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3 : \; (\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}}, f_{\mathbf{u},\mathbf{B_u},\mathbf{y}}(\mathbf{x}_i, \mathbf{x}_j) \geq f_{\mathbf{u},\mathbf{B_u},\mathbf{y}}(\mathbf{x}_1, \mathbf{x}_2)\right\}.
\end{aligned} \tag{72}$$

Next by symmetry we have

$$\begin{aligned}&\mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3:\ (\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}, f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_i,\mathbf{x}_j) \geq f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_1,\mathbf{x}_2)\right\}\\ &= \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3:\ (\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}, f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_i,\mathbf{x}_j) \leq f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_1,\mathbf{x}_2)\right\}. \qquad (73)\end{aligned}$$

To realize this, let $\mathbf{B}_\mathbf{u} = \{\mathbf{x}_m(\mathbf{u})\}_{m=1}^M$ be a codebook and let $i \geq$, $j \geq 3$, be a pair of indices such that $(\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}$, and $f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_i,\mathbf{x}_j) \leq f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_1,\mathbf{x}_2)$. Let π be a permutation such that $(\pi\mathbf{x}_i, \pi\mathbf{x}_j, \pi\mathbf{u}, \pi\mathbf{y}) = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{u}, \mathbf{y})$ (such a permutation exists since $(\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}$). Consider the codebook $\mathbf{B}'_\mathbf{u} = \left\{\mathbf{x}_1(\mathbf{u}), \mathbf{x}_2(\mathbf{u}), \mathbf{x}_i(\mathbf{u}), \mathbf{x}_j(\mathbf{u}), \{\pi\mathbf{x}_m(\mathbf{u})\}_{m\neq\{1,2,i,j\}}\right\}$. Obviously, by definition of π, we have $f_{\mathbf{u},\mathbf{B}'_\mathbf{u},\mathbf{y}}(\mathbf{x}_i,\mathbf{x}_j) \geq f_{\mathbf{u},\mathbf{B}'_\mathbf{u},\mathbf{y}}(\mathbf{x}_1,\mathbf{x}_2)$ and since

$$\mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\{\mathbf{B}_\mathbf{u}\} = \mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\{\mathbf{B}'_\mathbf{u}\} \qquad (74)$$

the symmetry argument holds.

Eqs. (73) yields

$$\begin{aligned}&\mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3:\ (\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}, f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_i,\mathbf{x}_j) \geq f_{\mathbf{u},\mathbf{B}_\mathbf{u},\mathbf{y}}(\mathbf{x}_1,\mathbf{x}_2)\right\}\\ &\geq \frac{1}{2}\cdot\mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3:\ (\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}\right\}. \qquad (75)\end{aligned}$$

Hence, (72) and (75) yield the lower bound

$$\begin{aligned}&\Pr\{\text{error } |\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}\}\\ &\geq \frac{1}{2}\cdot\mathbf{P}^M_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}\left\{\exists i \geq 3, j \geq 3:\ (\mathbf{x}_i, \mathbf{x}_j) \in T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{uy}}\right\}\\ &= P_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\\ &\geq q_{M-4}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\cdot\frac{C_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})}{1+C_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})}\\ &\geq q_{M-4}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\cdot\frac{1}{2}\min\{1, C_{M-2}(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\}\\ &\doteq q_M(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\cdot\frac{1}{2}\min\{1, C_M(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y})\}, \qquad (76)\end{aligned}$$

where the first inequality follows from (38), and the second inequality follows because either $C_M \leq 1$ and then $\frac{C_M}{1+C_M} \geq \frac{C_M}{2}$ or $C_M \geq 1$, and consequently $\frac{C_M}{1+C_M} \geq \frac{1}{2}$. Now, recall that $q_M(\mathbf{u}\mathbf{x}_1\mathbf{x}_2\mathbf{y}) = [1-a]^M$ where $a \doteq \frac{\max\left\{|T_{\mathbf{x}_1|\mathbf{u}\mathbf{x}_2\mathbf{y}}|,|T_{\mathbf{x}_2|\mathbf{u}\mathbf{x}_1\mathbf{y}}|\right\}}{|T_{\mathbf{x}|\mathbf{u}}|} \doteq e^{-n\min\left\{\hat{I}_{\mathbf{x}_1;\mathbf{x}_2\mathbf{y}|\mathbf{u}},\hat{I}_{\mathbf{x}_2;\mathbf{x}_1\mathbf{y}|\mathbf{u}}\right\}}$, so whenever $R \leq \min\left\{\hat{I}_{\mathbf{x}_1;\mathbf{x}_2\mathbf{y}|\mathbf{u}}, \hat{I}_{\mathbf{x}_2;\mathbf{x}_1\mathbf{y}|\mathbf{u}}\right\}$, we have $q_M \doteq e^{-1} \doteq 1$. Hence, using an equivalent of Lemma 3, the expression for the error exponent resulting from (76), is upper bounded (on the exponential scale) by

$$\begin{aligned}&\min_{\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}\in\mathcal{P}^{(n)}_{d_2}(\eta,D_2), R\leq\min\left\{\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}},\hat{I}_{\tilde{\mathbf{x}};\mathbf{x}\mathbf{y}|\mathbf{u}}\right\}}\\ &\left[D(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}||P_U\eta(\hat{P}_\mathbf{u})\hat{P}_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}) + \left|\hat{I}_{\tilde{\mathbf{x}};\mathbf{y}|\mathbf{u}} + \hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}} - 2R\right|^+\right], \qquad (77)\end{aligned}$$

which differs from $\tilde{E}_b(P_U, D_2, \eta, R)$ (see (22)) only by the fact that the minimization is over empirical measures of order n rather than over the continuum. Since η is a continuous mapping, the above expression converges to $\tilde{E}_b(P_U, D_2, \eta, R)$. This concludes the proof of the r.h.s. of (23).

Now, by definition of $R_0(\eta, D_2)$ (following e.q. (22)), $E_b(P_U, D_2, \eta, R) = \tilde{E}_b(P_U, D_2, \eta, R)$ for all $R \in [0, R_0(\eta, D_2)]$ and this concludes the proof of Theorem 2.

8 Proofs of Lemmas

Proof of Lemma 1

Proof. The r.h.s of (27) follows trivially. As for the l.h.s., it is easy to show that for all $a \in [0, 1]$,

$$1 - (1-a)^M \geq \frac{Ma}{1+Ma}. \tag{78}$$

To see that, note that if $(1-a)^M \leq \frac{1}{1+Ma}$, we have $1-(1-a)^M \geq \frac{Ma}{1+Ma}$, and if $(1-a)^M \geq \frac{1}{1+Ma}$, we have $1-(1-a)^M \geq aM(1-a)^{M-1} \geq \frac{a \cdot M}{(1-a)(1+Ma)}$, thus, $1-(1-a)^M \geq \min\left\{\frac{Ma}{1+Ma}, \frac{M}{(1-a)(1+Ma)}\right\} = \frac{Ma}{1+Ma}$.

The lemma follows since

$$\frac{Ma}{1+Ma} \geq \frac{1}{2}\min\{1, Ma\}. \tag{79}$$

□

Proof of Lemma 2

Proof. The first part of the lemma (37) follows trivially because $\left(\begin{array}{c} M \\ 2 \end{array}\right)$ is the number of possibilities of choosing a pair among the M vectors, and $\frac{|T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}}|}{|T_{\mathbf{x}|\mathbf{u}}|^2}$ is the probability that two i.i.d. vectors uniformly distributed over $T_{\mathbf{x}|\mathbf{u}}$ lie within $T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}}$.

In order to prove (38), note that

$$\begin{aligned}
&P_M \geq \mathbf{P}_{\mathbf{u}}^M \left\{ \text{ There exists a } \textit{single} \text{ pair } (i,j) \text{ s.t. } (\mathbf{X}_i, \mathbf{X}_j) \in T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}} \right\} \\
&= C_M \cdot \mathbf{P}_{\mathbf{u}}^M \left\{ (\mathbf{X}_1, \mathbf{X}_2) \text{ is the single pair in } T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}} \middle| (\mathbf{X}_1, \mathbf{X}_2) \in T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}} \right\} \\
&\overset{(a)}{=} C_M \cdot \mathbf{P}_{\mathbf{u}}^M \left\{ \mathbf{X}_i \notin \{T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}} \cup T_{\tilde{\mathbf{x}}|\mathbf{uxy}}\} \forall i \geq 3,\ (\mathbf{X}_i, \mathbf{X}_j) \notin T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}} \forall i,j \geq 3 \right\} \\
&\overset{(b)}{\geq} C_M \left[\mathbf{P}_{\mathbf{u}}^M \left\{ \mathbf{X}_i \notin \{T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}} \cup T_{\tilde{\mathbf{x}}|\mathbf{uxy}}\} \forall i \geq 3 \right\} + \mathbf{P}_{\mathbf{u}}^M \left\{ (\mathbf{X}_i, \mathbf{X}_j) \notin T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{uy}} \forall i,j \geq 3 \right\} - 1 \right] \\
&= C_M \left[q_{M-2} + 1 - P_{M-2} - 1 \right] \\
&\geq C_M \left[q_{M-2} - P_M \right]
\end{aligned} \tag{80}$$

where (a) is because if, for example, there exists $i \geq 3$ such that $\mathbf{X}_i \in T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}}$, then $T_{\mathbf{x}_i,\mathbf{x}_2|\mathbf{u}\mathbf{y}} = T_{\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}\mathbf{y}} = T_{\mathbf{x},\tilde{\mathbf{x}}|\mathbf{u}\mathbf{y}}$, (b) follows since for two events A, B, $1 \geq P(A \cup B) \geq P(A) + P(B) - P(A \cap B)$, thus, $P(A \cap B) \geq P(A) + P(B) - 1$, and the last step holds since by definition P_M is non decreasing with M. □

Proof of Lemma 3

Proof. To establish the proof we shall show that the r.h.s. of (48) is upper bounded (in the exponential scale) by the r.h.s. of (50), and that the r.h.s. of (61) is lower bounded by the r.h.s. of (50).

Let $P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ be a given attack channel, which is a member of $\mathcal{P}_n^{d_2}$, and let π be one of the $n!$ permutations of $\{1, ..., n\}$. Denote by $P^{\pi}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ the channel defined by

$$P^{\pi}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}) = P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\pi\mathbf{y}|\pi\mathbf{x}\pi\tilde{\mathbf{x}}), \tag{81}$$

where $\pi\mathbf{x}$ designates the sequence $\mathbf{x}$ permuted according to π.

For a given watermarking channel $P_{\mathbf{X}|\mathbf{U}} \in \mathcal{P}_n^{d_1}$ denote

$$L(P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}) \triangleq \sum_{\mathbf{u}} P_U^n(\mathbf{u}) \sum_{\mathbf{x},\tilde{\mathbf{x}} \in T_\eta(\mathbf{u})} \frac{1}{|T_\eta(\mathbf{u})|^2} \sum_{\mathbf{y}} P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}) \alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}). \tag{82}$$

Since P_U^n is memoryless, the encoder satisfies (32), $\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})$ depends solely on the type class $T_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}$, and since $L(P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}})$ is an affine functional of $P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$, we have

$$L\left(P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\right) = L\left(P^{\pi}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\right) = L\left(\frac{1}{n}\sum_{\pi} P^{\pi}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\right). \tag{83}$$

Note that $\frac{1}{n}\sum_{\pi} P^{\pi}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ is a strongly exchangeable channel, as if $T_{\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}} = T_{\mathbf{x}'\tilde{\mathbf{x}}'\mathbf{y}'}$, there exists a permutation $\bar{\pi}$ such that $(\mathbf{x}'\tilde{\mathbf{x}}'\mathbf{y}') = (\bar{\pi}\mathbf{x}\bar{\pi}\tilde{\mathbf{x}}\bar{\pi}\tilde{\mathbf{y}})$, and thus $\sum_{\pi} P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\pi\mathbf{y}|\pi\mathbf{x}\tilde{\mathbf{x}}) = \sum_{\pi} P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\pi\mathbf{y}'|\pi\mathbf{x}'\tilde{\mathbf{x}'})$.

Thus, we can denote that $\frac{1}{n}\sum_{\pi} P^{\pi}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}) = \frac{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|\mathbf{x}\tilde{\mathbf{x}})}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}$, and (83) implies

$$L\left(P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\right) = L\left(\frac{P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\left(T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|\mathbf{x}\tilde{\mathbf{x}}\right)}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}\right). \tag{84}$$

Now, observe that (10) implies

$$P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}(T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|\mathbf{x}\tilde{\mathbf{x}}) \leq \mathbf{1}\left\{\max\{d_2(\mathbf{x},\mathbf{y}), d_2(\tilde{\mathbf{x}},\mathbf{y})\} \leq nD_2\right\}, \tag{85}$$

and since $L(P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}})$ is affine in $P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ we have

$$L\left(P_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\right) \stackrel{\cdot}{\leq} L\left(P^{*}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}\right), \tag{86}$$

where $P^{*}_{\mathbf{Y}|\mathbf{X}\tilde{\mathbf{X}}}$ is defined in (26). Hence, the r.h.s. of (48) is upper bounded (in the exponential scale) by the r.h.s. of (50).

□

Proof of Lemma 4

First note that

$$
\begin{aligned}
&\sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}} \Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}) \sum_{\mathbf{y}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}}|} \\
= &\sum_{\mathbf{u}} P_U^n(\mathbf{u}) \sum_{\mathbf{x},\tilde{\mathbf{x}}\in T_\eta(\mathbf{u})} \frac{1}{|T_\eta(\mathbf{u})|^2} \sum_{\mathbf{y}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|} \\
= &\sum_{\mathbf{u}} P_U^n(\mathbf{u}) \sum_{\mathbf{x}\in T_\eta(\mathbf{u})} \frac{1}{|T_\eta(\mathbf{u})|} \sum_{T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}}\subseteq T_\eta(\mathbf{u})} \frac{|T_{\tilde{\mathbf{x}}|\mathbf{x}\mathbf{u}}|}{|T_\eta(\mathbf{u})|} \\
&\times \sum_{T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{|T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}|}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}). \qquad (87)
\end{aligned}
$$

Taking the minimum over $\eta\in\mathcal{N}_n(D_1)$, one notices that the minimizer satisfies $\eta(\mathbf{u}')=\eta(\mathbf{u})$ whenever $\mathbf{u}'\in T_{\mathbf{u}}$, because the quantity

$$
\sum_{T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}}\subseteq T_\eta(\mathbf{u})} \frac{|T_{\tilde{\mathbf{x}}|\mathbf{x}\mathbf{u}}|}{|T_\eta(\mathbf{u})|} \sum_{T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{|T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}|}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})
$$

depends only on the joint type class of $(\mathbf{u},\mathbf{x})$ thus the minimizing $\eta\in\mathcal{N}_n(D_1)$ belongs to $\mathcal{N}_n^{ex}(D_1)$ which proves (51). Next note that since $\eta\in\mathcal{N}_n(D_1)$ is a mapping from $\mathbf{u}$ to $\mathcal{P}(\mathcal{X}|\mathcal{U})$, one can switch the order between the summation over $\mathbf{u}\in\mathcal{U}^n$ and the minimization over $\eta\in\mathcal{N}_n(D_1)$, i.e.,

$$
\begin{aligned}
&\min_{\eta\in\mathcal{N}_n(D_1)} \sum_{\mathbf{u}} P_U^n(\mathbf{u}) \sum_{\mathbf{x}\in T_\eta(\mathbf{u})} \frac{1}{|T_\eta(\mathbf{u})|} \sum_{T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}}\subseteq T_\eta(\mathbf{u})} \frac{|T_{\tilde{\mathbf{x}}|\mathbf{x}\mathbf{u}}|}{|T_\eta(\mathbf{u})|} \\
&\times \sum_{T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{|T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}|}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) \\
= &\sum_{\mathbf{u}} P_U^n(\mathbf{u}) \min_{\eta(\mathbf{u}):\,d_1(\mathbf{u},\mathbf{x})\le nD_1} \sum_{\mathbf{x}\in T_\eta(\mathbf{u})} \frac{1}{|T_\eta(\mathbf{u})|} \sum_{T_{\tilde{\mathbf{x}}|\mathbf{u}\mathbf{x}}\subseteq T_\eta(\mathbf{u})} \frac{|T_{\tilde{\mathbf{x}}|\mathbf{x}\mathbf{u}}|}{|T_\eta(\mathbf{u})|} \\
&\times \sum_{T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{|T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}|}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}). \qquad (88)
\end{aligned}
$$

Now we use the method of types to evaluate the r.h.s. expression in the exponential scale

$$\begin{aligned}
&\sum_{\mathbf{u}} P_U^n(\mathbf{u}) \min_{\eta(\mathbf{u}):\, d_1(\mathbf{u},\mathbf{x})\le nD_1} \sum_{\mathbf{x}\in T_\eta(\mathbf{u})} \frac{1}{|T_\eta(\mathbf{u})|} \sum_{T_{\tilde{\mathbf{x}}|\mathbf{ux}}\subseteq T_\eta(\mathbf{u})} \frac{|T_{\tilde{\mathbf{x}}|\mathbf{xu}}|}{|T_\eta(\mathbf{u})|} \\
&\times \sum_{T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}:\, \max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} \frac{|T_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}|}{|T_{\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}|}\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) \qquad (89)\\
\doteq\ & \max_{\hat{P}_{\mathbf{u}}} e^{-nD(\hat{P}_{\mathbf{u}}||P_U)} \min_{\eta(\mathbf{u}):\, d_1(\mathbf{u},\mathbf{x})\le nD_1} \max_{T_{\tilde{\mathbf{x}}|\mathbf{ux}}\subseteq T_\eta(\mathbf{u})} e^{-n\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}|\mathbf{u}}} \\
&\times \max_{\hat{P}_{\mathbf{y}|\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}}:\, \max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2} e^{-n\hat{I}_{\mathbf{u};\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}}\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) \\
=\ & \exp\left\{-n\left(\min\max\min\left[D(\hat{P}_{\mathbf{u}}||P_U)+\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}|\mathbf{u}}+\hat{I}_{\mathbf{u};\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}}-\frac{1}{n}\log\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})\right]\right)\right\} \qquad (90)
\end{aligned}$$

where the outmost minimization is over $\hat{P}_{\mathbf{u}} \in \mathbb{P}_n(\mathcal{U})$, the maximization is over $\hat{P}_{\mathbf{x}|\mathbf{u}} \in \mathbb{P}_n(\mathcal{X}, \hat{P}_{\mathbf{u}})$ and the inner minimization is over $\hat{P}_{\tilde{\mathbf{x}}\mathbf{y}|\mathbf{ux}} \in \mathbb{P}_n(\mathcal{X}\times\mathcal{Y}, \hat{P}_{\mathbf{ux}})$ such that the marginal distributions satisfy $\hat{P}_{\tilde{\mathbf{x}}|\mathbf{u}} = \hat{P}_{\mathbf{x}|\mathbf{u}}$ and $\max\{d_1(\mathbf{x},\mathbf{y}), d_1(\tilde{\mathbf{x}},\mathbf{y})\} \le nD_2$. The last step follows from the fact that $\hat{P}_{\mathbf{x}|\mathbf{u}} = \hat{P}_{\tilde{\mathbf{x}}|\mathbf{u}} = \eta(\hat{P}_{\mathbf{u}})$.

From (19), we have,

$$\begin{aligned}
D(\hat{P}_{\mathbf{u}}||P_U)+\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}|\mathbf{u}}+\hat{I}_{\mathbf{u};\mathbf{y}|\mathbf{x}\tilde{\mathbf{x}}} &= D(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}||P_U\times\hat{P}_{\mathbf{x}|\mathbf{u}}\times\hat{P}_{\tilde{\mathbf{x}}|\mathbf{u}}\times\hat{P}_{\hat{\mathbf{y}}|\mathbf{x}\tilde{\mathbf{x}}}) \\
&= D(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}, P_U). \qquad (91)
\end{aligned}$$

Recall the definition of a (44), and note that $a \doteq \frac{\max\left\{|T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}}|,|T_{\mathbf{x}|\mathbf{u}\tilde{\mathbf{x}}\mathbf{y}}|\right\}}{|T_\eta(\mathbf{u})|} \doteq e^{-n\min\left\{\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}},\hat{I}_{\tilde{\mathbf{x}};\mathbf{x}\mathbf{y}|\mathbf{u}}\right\}}$. We also have by (40), $C_M \doteq e^{n\left(2R-\hat{I}_{\tilde{\mathbf{x}};\mathbf{y}|\mathbf{u}}-\hat{I}_{\mathbf{x};\mathbf{y}|\mathbf{u}\tilde{\mathbf{x}}}\right)}$. Hence, by (47), we have,

$$\begin{aligned}
-\frac{1}{n}\log\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}}) &= -\frac{1}{n}\log\min\{1,\max\{Ma\ ,\ C_{M-2}\}\} \\
\doteq\ & \left|\min\left\{\min\left\{\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}},\hat{I}_{\tilde{\mathbf{x}};\mathbf{x}\mathbf{y}|\mathbf{u}}\right\}-R\ ,\ \hat{I}_{\tilde{\mathbf{x}};\mathbf{y}|\mathbf{u}}+\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}}-2R\ \right\}\right|^+ \\
=\ & \left|\min\left\{\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}}-R\ ,\ \hat{I}_{\tilde{\mathbf{x}};\mathbf{x}\mathbf{y}|\mathbf{u}}-R\ ,\ \hat{I}_{\tilde{\mathbf{x}};\mathbf{y}|\mathbf{u}}+\hat{I}_{\mathbf{x};\tilde{\mathbf{x}}\mathbf{y}|\mathbf{u}}-2R\right\}\right|^+. \qquad (92)
\end{aligned}$$

Thus (88)-(92) imply

$$\begin{aligned}
&\min_{\eta\in\mathcal{N}_n(D_1)}\sum_{\mathbf{u},\mathbf{x},\tilde{\mathbf{x}}}\Pr(\mathbf{u},\mathbf{x},\tilde{\mathbf{x}})\sum_{\mathbf{y}:\,\max\{d_2(\mathbf{x},\mathbf{y}),d_2(\tilde{\mathbf{x}},\mathbf{y})\}\le nD_2}\frac{\alpha_M(\hat{P}_{\mathbf{u}\mathbf{x}\tilde{\mathbf{x}}\mathbf{y}})}{|T_{\mathbf{y}|\mathbf{x},\tilde{\mathbf{x}}}|} \\
\doteq\ & \exp\left(-n\min_{P_{\tilde{U}}}\max_{P_{X|\tilde{U}}}\min_{P_{\tilde{X}Y|\tilde{U},X}}\left[D(P_{\tilde{U}X\tilde{X}Y},P_U)+\epsilon_a(P_{\tilde{U}X\tilde{X}Y},R)\right]\right), \qquad (93)
\end{aligned}$$

where the outmost minimization is over $P_{\tilde{U}} \in \mathbb{P}_n(\mathcal{U})$, the maximization is over $P_{X|\tilde{U}} \in \mathbb{P}_n(\mathcal{X}, P_{\tilde{U}}) :\ Ed_1(\tilde{U}, X) \le D_1$, and the innermost minimization is over $P_{\tilde{X}Y|\tilde{U},X} \in \mathbb{P}_n(\mathcal{X}\times\mathcal{Y}, P_{\tilde{U},X})$: s.t. $P_{\tilde{X}|\tilde{U}} = P_{X|\tilde{U}}$ and $\max\{Ed_2(X,Y), Ed_2(\tilde{X},Y)\} \le D_2$.

References

[1] A. Barg, G. R. Blakley, and G. Kabatiansky, "Good digital fingerprinting codes," *Proc. ISIT 2001*, p. 161, Washington, D.C., June 2001.

[2] A. Barg, G. R. Blakley, and G. Kabatiansky, "Digital fingerprinting codes: problem statements, constructions, identification of traitors," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 852-865, Apr. 2003.

[3] D. Boneh and J. Shaw, "Collusion–secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1897–1905, September 1998.

[4] B. Chen and G.W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[5] B. Chor, A. Fiat, M. Naor, and B. Pinkas, "Tracing traitors," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 893–910, May 2000.

[6] G. Cohen, S. Encheva, and G. Zémor, "Copyright protection for digital data," *IEEE Communication Letters*, vol. 4, no. 5, pp. 158–160, May 2000.

[7] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639-1667, June 2002.

[8] A. Cohen and A. Lapidoth, "The capacity of the vector Gaussian watermarking game," in *Proc. ISIT 2001*, p. 5, 2001.

[9] G.D. Cohen and H.G. Schaathun, "Upper bounds on separating codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1291–1294, June 2004.

[10] T. M. Cover, "Enumerative source coding," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 1, pp. 73–77, Jan. 1973.

[11] I. Csiszár and J. Körner, *Coding theorems for discrete memoryless systems.* New York: Academic, 1981.

[12] J. Dittmann, A. Behr, M. Stabenau, P. Schmitt, J. Schwenk, and J. Ueberberg, Combining digital watermarks and collusion secure fingerprints for digital images," preprint, 2001.

[13] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 244-246, Mar. 1973.

[14] D. Karakos, and A. Papamarcou, "A relationship between quantization and watermarking rates in the presence of additive Gaussian attacks Information Theory," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1970–1982, Aug. 2003.

[15] T. Lindkvist, J. Lofvenberg, and M. Svanstrom, "A class of traceability codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 2094–2096 July 2002.

[16] Y. S. Liu and B. L. Hughes, "A new universal random coding bound for the multiple access channel," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.

[17] J. Löfvenberg, *Codes for digital fingerprints*, Ph.D. dissertation, Department of Electrical Engineering, Linköping University, Sweden, June 2001.

[18] J. Löfvenberg and N. Wilberg, "Random codes for digital fingerprinting," Report Reg nr:LiTH-ISY-R-2059, Department of Electrical Engineering, Linköping University, Sweden, June 2001.

[19] N. Merhav, "On random coding error exponents of watermarking systems," *IEEE Trans. Inform. Theory*, vol. IT–46, no. 2, pp. 420-430, March 2000.

[20] P. Moulin and A. Briassouli, "The Gaussian fingerprinting game," *Proc. 2002 Conference on Information Sciences and Systems*, Princeton University, March 20–22, 2002.

[21] P. Moulin and M.K. Mihcak, "The parallel-Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, pp. 272–289, Feb. 2004.

[22] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," in *Proc. ISIT 2000*, p. 19, Sorrento, Italy, June 2000.

[23] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, March 2003.

[24] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, "Information–theoretic analysis of steganography," in *Proc. ISIT '98*, M.I.T., Cambridge, MA, p. 297, August 1998.

[25] S.S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.

[26] R. Safavi–Naini and Y. Wang, "Collusion secure q–ary fingerprinting for perceptual content," preprint 2001.

[27] R. Safavi–Naini and Y. Wang, "Sequential traitor tracing," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1319-1326, May 2003.

[28] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel Aviv University, 2003.

[29] A. Silverberg, J. Staddon, and J.L. Walker, "Application of list decoding to tracing traitors," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1312-1318, May 2003.

[30] A. Somekh-Baruch and N. Merhav, "On the error exponent and capacity games of private watermarking systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 537-562, March 2003.

[31] A. Somekh-Baruch and N. Merhav, "On the capacity game of public watermarking systems," *IEEE Trans. Inform. Theory*, vol. 50, no. 3, pp. 511-524, March 2004.

[32] A. Somekh–Baruch and N. Merhav, "On the capacity game of fingerprinting systems under collusion attacks," to appear in *IEEE Trans. Inform. Theory*, March 2004.

[33] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1410-1422, May 2001.

[34] J. K. Su, J. J. Eggers, and B. Girod, "Capacity of digital watermarks subjected to an optimal collusion attack," *Proc. EUSIPCO 2000.*

[35] K. Su, D. Kundur, and D. Hatzinakos, "A novel approach to collusion–resistant video watermarking," preprint, 2001.

[36] C. Xing "Asymptotic bounds on frameproof codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2991–2995, Nov. 2004.