IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

False–Accept/False–Reject Trade–offs in Biometric Authentication Systems

Neri Merhav

CCIT Report #924
May 2018

THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 3200003, ISRAEL

CCIT Report #924 May 2018

False–Accept/False–Reject Trade–offs in Biometric Authentication Systems

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

Biometric authentication systems, based on secret key generation, work as follows. In the enrollment stage, an individual provides a biometric signal that is mapped into a secret key and a helper message, the former being prepared to become available to the system at a later time (for authentication), and the latter is stored in a public database. When an authorized user requests authentication, claiming his/her identity as one of the subscribers, he/she has to provide a biometric signal again, and then the system, which retrieves also the helper message of the claimed subscriber, produces an estimate of the secret key, that is finally compared to the secret key of the claimed user. In case of a match, the authentication request is approved, otherwise, it is rejected.

Evidently, there is an inherent tension between two desired, but conflicting, properties of the helper message encoder: on the one hand, the encoding should be informative enough concerning the identity of the real subscriber, in order to approve him/her in the authentication stage, but on the other hand, it should not be too informative, as otherwise, unauthorized imposters could easily fool the system and gain access. A good encoder should then trade off the two kinds of errors: the false reject (FR) error and the false accept (FA) error.

In this work, we investigate trade–offs between the random coding FR error exponent and the best achievable FA error exponent. We compare two types of ensembles of codes: fixed–rate codes and variable–rate codes, and we show that the latter class of codes offers considerable improvement compared to the former. In doing this, we characterize the optimal rate functions for both types of codes. We also examine the effect of privacy leakage constraints on these trade-offs, for both fixed–rate codes and variable–rate codes.

Index Terms: biometric systems, secret sharing, error exponents, random binning, fixed–length, variable–length, privacy leakage.

1 Introduction

We consider a biometric authentication system which is based on the one described in [6, Sections 2.2–2.6], and on the notion of secret key generation and sharing of Maurer [7] and Ahlswede and Csiszár [1], [2]. In particular, this system works in the following manner. In the enrollment phase, a person that subscribes to the system, feeds it with a biometric signal, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The system then responds by generating (using its encoder) two outputs. The first is a secret key, $\boldsymbol{S}$, at rate $R_{\rm s}$ and the second is a helper message, $\boldsymbol{W}$, at rate $R_{\rm w}$. The secret key will be used by the system later, at the authentication stage and the helper message is saved in a database. When an authorized user (a subscriber) wishes to sign in, claiming to be one of the subscribers that have already enrolled, he/she is requested to provide again his/her biometric signal, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ (correlated to $\boldsymbol{X}$, if indeed from the same person, or independent, otherwise). The system then retrieves the helper message $\boldsymbol{W}$ of the claimed subscriber from the database, and responds (using its decoder) by estimating the secret key, $\hat{\boldsymbol{S}}$ (based on $(\boldsymbol{Y}, \boldsymbol{W})$), and comparing it to that of the claimed user, $\boldsymbol{S}$. If $\hat{\boldsymbol{S}}$ matches $\boldsymbol{S}$, the access to the system is approved, otherwise, it is denied.

In [6, Sect. 2.3], the achievable region of pairs of rates $(R_{\rm s}, R_{\rm w})$ was established for the existence of authentication systems where the following four quantities need to be arbitrarily small for large n: (i) the false–reject (FR) probability, (ii) the false–accept (FA) probability, (iii) the privacy leakage, $I(\boldsymbol{X};\boldsymbol{W})/n$, and (iv) the secrecy leakage, $I(\boldsymbol{S};\boldsymbol{W})/n$. Specifically, Theorem 2.1 of [6] asserts that when $(\boldsymbol{X}, \boldsymbol{Y})$ are drawn from a joint discrete memoryless source (DMS), emitting independent copies of a pair of dependent random variables, $(X, Y) \sim P_{XY}$, the largest achievable key rate, $R_{\rm s}$, under the above constraints, is given by $I(X;Y)$. It then follows that $R_{\rm w}$ must lie between $H(X|Y)$ and $H(X) - R_{\rm s}$, where the lower limit is needed for good identification of the legitimate subscriber (small FR probability) as well as for achieving the minimum possible privacy leakage, whereas the upper limit is due to the secrecy leakage requirement. These limitations already assure that $R_{\rm w} < H(X)$, which in turn is necessary for keeping the FA probability vanishingly small for large n. The achievability parts of the corresponding coding theorems were proved in [6] using random binning, similarly as in classical Slepian–Wolf coding.

More recently, in [9] these results have been refined by characterizing achievable exponential error bounds for the above performance metrics. In particular, for a given rate pair $(R_{\rm s}, R_{\rm w})$,

random coding error exponents and expurgated error exponents were found for the FR probability, as well as sphere–packing bound, which is tight at a certain region of the plane of $(R_{\rm s}, R_{\rm w})$. For the FA probability, the exact best achievable error exponent was characterized, and finally, more refined upper bounds for privacy leakage and the secrecy leakage were derived.

This paper is a further development of [9], where the focus is on the trade–off between the FA error exponent and the FR error exponent. In the design of the helper message encoder, the following conflict arises: on the one hand, it is desirable that the helper message $\boldsymbol{W}$ would be informative enough about $\boldsymbol{S}$, such that in the presence of $\boldsymbol{Y}$, the identity of the legitimate subscriber will be approved with high probability. But on the other hand, it is also desired that in the absence of $\boldsymbol{Y}$, the helper message would tell as little as possible about $\boldsymbol{S}$, in order to make it difficult for imposters to access the system.

Indeed, the converse theorem in [9, Theorem 5] is based on the assumption that every type class of source sequences $\{\boldsymbol{X}\}$, is mapped, by the helper–message encoder, to as many different helper messages $\{\boldsymbol{W}\}$ as possible, thus making it as close as possible to be a one–to–one mapping, or in other words, making $\boldsymbol{W}$ is "as informative as possible" about the source vector $\boldsymbol{X}$, and hence also about the secret key $\boldsymbol{S}$, generated from $\boldsymbol{X}$. This is good for achieving a small FR probability (or, equivalently, a large FR error exponent), at the expense of a limitation on the achievable FA exponent. In particular, by relaxing the above described assumption, and allowing smaller numbers of various helper messages for each source type class, one may achieve better FA exponents, at the expense of worse FR exponents.

This raises the interesting question of achievable trade-offs between the FA exponent and the FR exponent, which is similar, in spirit, to the trade–off between the false alarm probability and the mis–detection probability in the Neyman–Pearson scenario, where this trade–off is traditionally encapsulated by the notion of receiver operating characteristics (ROC). The difference, however, is that while in the Neyman–Pearson setting, this trade–off is controlled by the choice of a detector (or more precisely, by the choice of the threshold of the likelihood ratio detector), here we control the trade-off via the choice of an encoder, in this case, the helper–message encoder.

To this end, we first derive an expression of the FR random coding error exponent as a function of the desired FA error exponent for fixed–rate binning. This is a relatively straightforward

manipulation of the results of [9], but it will serve as a reference result. The more interesting part, however, is about extending the scope to the ensemble of variable–rate random binning codes, whose *rate function* depends on the source vector only via its type (similarly as in [12] and [3]). The are two questions that arise in this context. The first is: **what are the optimal rate functions** of the secret–message and the helper–message for maximizing the FR exponent for a given FA error exponent? Upon finding these optimal rate functions, the second question is: what is the achievable FR error exponent as a function of the FA error exponent, and to what extent does it improve relative to fixed–rate binning? We find an exact formula for this function and demonstrate that the improvement may be rather significant compared to fixed–rate binning.

Finally, we examine the influence of adding a constraint on the privacy leakage, in addition to the above mentioned constraint on the FA error exponent. We show that under certain conditions, the privacy leakage constraint causes some deterioration in performance for fixed–rate codes, but not in variable–rate codes, thus increasing the gap between variable–rate codes and fixed–rate codes even further.

On a technical note, it should be pointed out that while in [9], the error exponent expressions are provided in the Csiszár–style formulation (i.e., minimizations of certain functionals of information measures over probability distributions), here we pass to Gallager–style formulations (i.e., maximizations of functions of relatively few parameters). The reasons for our interest in Gallager–style expressions are that they lend themselves more conveniently to numerical calculations (see the discussion after Theorem 1 below, for more details), and that they may be of independent interest on their own right.

The outline of the remaining part of this paper is as follows. Section II establishes the notation conventions. Section III provides a formal definition of the problem setting, then it gives some background (preliminaries), and finally, it describes the objectives. Section IV provides a preparatory step of deriving the optimal rate functions that maximize the the FR error exponent for a given FA error exponent, in both fixed–rate and variable–rate regimes. In Section V, we derive the FA error exponents of both fixed– and variable–rate codes as functions the prescribed FA error exponent. Finally, in Section VI, we examine the effect of a constraint on the privacy leakage.

II. Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the n–th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. Sources and channels will be denoted by the letter P or Q, subscripted by the names of the relevant random variables/vectors and their conditionings, if applicable, following the standard notation conventions, e.g., Q_X, $P_{Y|X}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. The probability of an event $\mathcal{G}$ will be denoted by $\Pr\{\mathcal{G}\}$, and the expectation operator with respect to (w.r.t.) a probability distribution P will be denoted by $\boldsymbol{E}_P\{\cdot\}$. Again, the subscript will be omitted if the underlying probability distribution is clear from the context. The entropy of a generic distribution Q on $\mathcal{X}$ will be denoted by $H_Q(X)$. For two positive sequences a_n and b_n, the notation $a_n \doteq b_n$ will stand for equality in the exponential scale, that is, $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. Similarly, $a_n \stackrel{\cdot}{\leq} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \leq 0$, and so on. The indicator function of an event $\mathcal{G}$ will be denoted by $\mathcal{I}\{\mathcal{G}\}$. The notation $[x]_+$ will stand for $\max\{0, x\}$.

The empirical distribution of a sequence $\boldsymbol{x} \in \mathcal{X}^n$, which will be denoted by $\hat{P}_{\boldsymbol{x}}$, is the vector of relative frequencies $\hat{P}_{\boldsymbol{x}}(x)$ of each symbol $x \in \mathcal{X}$ in $\boldsymbol{x}$. The type class of $\boldsymbol{x} \in \mathcal{X}^n$, denoted $\mathcal{T}(\hat{P}_{\boldsymbol{x}})$, is the set of all vectors $\boldsymbol{x}'$ with $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$. Information measures associated with empirical distributions will be denoted with 'hats' and will be subscripted by the sequences from which they are induced. For example, the entropy associated with $\hat{P}_{\boldsymbol{x}}$, which is the empirical entropy of $\boldsymbol{x}$, will be denoted by $\hat{H}_{\boldsymbol{x}}(X)$. Similar conventions will apply to the joint empirical distribution, the joint type class, the conditional empirical distributions and the conditional type classes associated with pairs (and multiples) of sequences of length n. Accordingly, $\hat{P}_{\boldsymbol{xy}}$ will be the joint empirical distribution of $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}_{i=1}^n$, and $\mathcal{T}(\hat{P}_{\boldsymbol{xy}})$ will denote the joint type class of $(\boldsymbol{x}, \boldsymbol{y})$. Similarly, $\mathcal{T}(\hat{P}_{\boldsymbol{x}|\boldsymbol{y}}|\boldsymbol{y})$ will stand for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X, Y)$ will designate the empirical joint entropy of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X|Y)$ will be the empirical conditional entropy, $\hat{I}_{\boldsymbol{xy}}(X; Y)$ will denote

empirical mutual information, and so on. We will also use similar rules of notation in the context of a generic distribution, Q_{XY} (or Q, for short): we use $\mathcal{T}(Q_X)$ for the type class of sequences with empirical distribution Q_X, $H_Q(X)$ – for the corresponding empirical entropy, $\mathcal{T}(Q_{XY})$ – for the joint type class x, $T(Q_{X|Y}|\boldsymbol{y})$ – for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $H_Q(X,Y)$ – for the joint empirical entropy, $H_Q(X|Y)$ – for the conditional empirical entropy, $I_Q(X;Y)$ – for the empirical mutual information, and so on. We will also use the customary notation for the weighted divergence,

$$D(Q_{Y|X}\|P_{Y|X}|Q_X) = \sum_{x\in\mathcal{X}} Q_X(x) \sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{P_{Y|X}(y|x)}. \tag{1}$$

III. Problem Setting, Preliminaries and Objectives

A. Problem Setting

The problem setting is similar to the one in [9], but with a few small differences, mainly related to the fact that here, in contrast to [9], we allow variable–rate binning codes.

Consider the following system model for biometric identification. An *enrollment source sequence*, $\boldsymbol{x} = (x_1, \ldots, x_n)$, that is a realization of the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, that emerges from a discrete memoryless source (DMS), P_X, with a finite alphabet $\mathcal{X}$, is fed into an *enrollment encoder*, $\mathcal{E}$, that generates two outputs: a secret key, $\boldsymbol{s}$ (a realization of a random variable $\boldsymbol{S}$), and a helper message, $\boldsymbol{w}$ (a realization of $\boldsymbol{W}$), both taking values in finite alphabets, $\mathcal{S}_n$ and $\mathcal{W}_n$, respectively. In the *fixed–rate regime*, $\mathcal{S}_n = \{0, 1, \ldots, e^{nR_{\rm s}}\}$ and $\mathcal{W}_n = \{0, 1, \ldots, e^{nR_{\rm w}}\}$, where $R_{\rm s}$ is the *secret–key rate*, and $R_{\rm w}$ is the *helper–message rate*. In the *variable–rate regime*, we allow both rates to depend on the type Q_X of the given input vector $\boldsymbol{x}$. In particular, in the variable rate regime, each type class $\mathcal{T}(Q_X)$, of source vectors of length n, is mapped, by the secret–key encoder and by the helper–message encoder, into $\mathcal{S}_n(Q_X) = \{0, 1, \ldots, e^{nR_{\rm s}(Q_X)}\}$ and $\mathcal{W}_n(Q_X) = \{0, 1, \ldots, e^{nR_{\rm w}(Q_X)}\}$, respectively, where $R_{\rm s}(Q_X)$ and $R_{\rm w}(Q_X)$, henceforth referred to as *rate functions*, are given continuous functions of Q_X. These encodings designate the enrollment stage.

Since the fixed–rate case is obviously a special case of the variable–rate case, our description will henceforth relate to the variable–rate case, with the understanding that in the fixed–rate case,

$R_{\rm s}(Q_X)$ and $R_{\rm w}(Q_X)$ are just constants, denoted $R_{\rm s}$ and $R_{\rm w}$, independent of Q_X.

As in [6], we consider the ensemble of enrollment encoders, $\{\mathcal{E}\}$, generated by *random binning*, where for each source vector $\boldsymbol{x} \in \mathcal{X}$, one selects independently at random, both a secret key and a helper message, under the uniform distributions across $\mathcal{S}_n(Q_X)$ and $\mathcal{W}_n(Q_X)$, respectively. We denote by $\boldsymbol{w} = f(\boldsymbol{x})$ and $\boldsymbol{s} = g(\boldsymbol{x})$, the randomly selected bin assignments for both outputs.

The *authentication decoder*, $\mathcal{A}$, which is aware of the randomly selected encoder, $\mathcal{E}$, is fed by two inputs: the helper message $\boldsymbol{w}$ and an *authentication source sequence*, $\boldsymbol{y} = (y_1, \ldots, y_n)$ (a realization of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$), that is produced at the output of a discrete memoryless channel (DMC), $P_{Y|X}$, with a finite output alphabet $\mathcal{Y}$, that is fed by $\boldsymbol{x}$. The output of the authentication decoder is $\hat{\boldsymbol{s}} = U(\boldsymbol{y}, \boldsymbol{w})$ (a realization of $\hat{\boldsymbol{S}}$), which is an estimate (possibly, randomized) of the secret key, $\boldsymbol{s}$. If $\hat{\boldsymbol{s}} = \boldsymbol{s}$, access to the system is granted, otherwise, it is denied. This decoding operation stands for the authentication stage.

The optimal estimator of $\boldsymbol{s}$, based on $(\boldsymbol{y}, \boldsymbol{w})$, in the sense of minimum FR probability, $\Pr\{\hat{\boldsymbol{S}} \neq \boldsymbol{S}\}$, is the maximum a posteriori probability (MAP) estimator, given by

$$\hat{\boldsymbol{s}}_{\rm MAP} = U(\boldsymbol{y}, \boldsymbol{w}) \stackrel{\Delta}{=} \arg\max_{\boldsymbol{s}} P(\boldsymbol{s}, \boldsymbol{w}|\boldsymbol{y}) = \arg\max_{\boldsymbol{s}} \sum_{\boldsymbol{x} \in \mathcal{X}^n} P(\boldsymbol{x}|\boldsymbol{y}) \cdot \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{w}\} \cdot \mathcal{I}\{g(\boldsymbol{x}) = \boldsymbol{s}\}, \quad (2)$$

where $P(\boldsymbol{x}|\boldsymbol{y})$ (shorthand notation for $P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$) is the posterior probability of $\boldsymbol{X} = \boldsymbol{x}$ given $\boldsymbol{Y} = \boldsymbol{y}$, that is induced by the product distribution, P_{XY} (and the subscript XY will sometimes be suppressed for simplicity, when there is no risk of compromising clarity).

As in [9], here too, we consider the framework of generalized stochastic likelihood decoders (GLDs) [8], [10], [11], [14], where the decoder randomly selects its output $\hat{\boldsymbol{s}}$ according to the posterior distribution

$$\tilde{P}(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{w}) = \frac{\sum_{\boldsymbol{x} \in \mathcal{X}^n} \exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} \cdot \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{w}\} \cdot \mathcal{I}\{g(\boldsymbol{x}) = \boldsymbol{s}\}}{\sum_{\boldsymbol{x} \in \mathcal{X}^n} \exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} \cdot \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{w}\}}, \quad (3)$$

where the function $a(\cdot)$, which will be referred to as the *decoding metric*, is any continuous function of the joint empirical distribution $\hat{P}_{\boldsymbol{x}\boldsymbol{y}}$. As explained in [9], as well as in earlier studies, the motivation for considering GLDs is that they provide a unified framework for examining a large variety of decoders. For example, with

$$a(\hat{P}_{\boldsymbol{x}\boldsymbol{y}}) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \hat{P}_{\boldsymbol{x}\boldsymbol{y}}(x, y) \ln P_{X|Y}(x|y), \quad (4)$$

we have the ordinary likelihood decoder [10], [11], [14]. For

$$a(\hat{P}_{\boldsymbol{xy}}) = \beta \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \hat{P}_{\boldsymbol{xy}}(x,y) \ln P_{X|Y}(x|y), \tag{5}$$

$\beta > 0$ being a parameter, we extend this to a parametric family of decoders. In particular, $\beta \to \infty$ leads to the ordinary MAP decoder, $\hat{\boldsymbol{s}}_{\rm MAP}$. Other choices of $a(\cdot)$ are associated with mismatched metrics,

$$a(\hat{P}_{\boldsymbol{xy}}) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \hat{P}_{\boldsymbol{xy}}(x,y) \ln P'(x|y), \tag{6}$$

P' being different from $P_{X|Y}$, and

$$a(\hat{P}_{\boldsymbol{xy}}) = -\beta \hat{H}_{\boldsymbol{xy}}(X|Y), \tag{7}$$

which for $\beta \to \infty$, tends to the universal minimum entropy decoder. When $a(\hat{P}_{\boldsymbol{xy}}) = \beta \cdot \alpha(\hat{P}_{\boldsymbol{xy}})$, $\alpha(\cdot)$ being an arbitrary function and $\beta \to \infty$, we end up with Csiszár's α–decoder [4].

An illegal user (imposter), who claims for a given legal identity, does not have the correlated biometric data $\boldsymbol{y}$, and so, the best he/she can do is to estimate $\boldsymbol{s}$ based on $\boldsymbol{w}$, and then forge any fake biometric data $\tilde{\boldsymbol{y}}$, which together with $\boldsymbol{w}$, would cause the decoder to output this estimate of $\boldsymbol{s}$. More precisely, the imposter first estimates $\boldsymbol{s}$ according to

$$\tilde{\boldsymbol{s}} = V(\boldsymbol{w}) \stackrel{\Delta}{=} \arg\max_{\boldsymbol{s}} P(\boldsymbol{s}|\boldsymbol{w}) = \arg\max_{\boldsymbol{s}} \sum_{\boldsymbol{x}\in\mathcal{X}^n} P(\boldsymbol{x}) \cdot \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{w}\} \cdot \mathcal{I}\{g(\boldsymbol{x}) = \boldsymbol{s}\}, \tag{8}$$

and then generates any $\tilde{\boldsymbol{y}} \in \mathcal{Y}^n$ such that $U(\tilde{\boldsymbol{y}}, \boldsymbol{w}) = \tilde{\boldsymbol{s}}$, and uses it as the biometric signal for authentication.

B. Preliminaries

In [9, Theorems 1, 4, and 5], the following two results (among others) were derived for fixed–rate binning at rates $R_{\rm w}$ and $R_{\rm s}$: the best achievable FA exponent is given by

$$E_{\rm FA}(R_{\rm w}, R_{\rm s}) = \min_{Q_X} [D(Q_X \| P_X) + \min\{R_{\rm s}, [H_Q(X) - R_{\rm w}]_+\}], \tag{9}$$

and the random coding FR exponent is given by,

$$E_{\rm FR}(R_{\rm w}) = \min_{Q_{X_0Y}} \{D(Q_{X_0Y} \| P_{XY}) + E(R_{\rm w}, Q_{X_0Y})\}, \tag{10}$$

where

$$E(R_{\rm w}, Q_{X_0Y}) = \min_{Q_{X|Y}} [R_{\rm w} - H_Q(X|Y) + [a(Q_{X_0Y}) - a(Q_{XY})]_+]_+. \tag{11}$$

As shown in [9, eq. (12)], for the decoding metric $a(Q) = -H_Q(X|Y)$, eq. (11) simplifies to

$$E(R_{\rm w}, Q_{X_0Y}) = [R_{\rm w} - H_Q(X_0|Y)]_+, \tag{12}$$

which is equivalent to the error exponent expression corresponding to optimal MAP decoding, $\hat{\boldsymbol{s}}_{\rm MAP}$ (i.e., eq. (5) with $\beta \to \infty$).

C. Objectives

As described in the Introduction, our first objective is to derive the FR error exponent as a function of the prescribed FA error exponent, henceforth referred to as the *FR–FA trade-off function*, for fixed–rate codes with optimal rate functions and optimal decoding metrics. This FR–FA trade-off function will be derived from eqs. (9)–(11). The more interesting goal would then be to extend the scope to variable–rate codes, derive the optimal rate functions, $R_{\rm w}^*(Q_X)$ and $R_{\rm s}^*(Q_X)$, then use them to obtain the FR–FA trade–off function for variable–rate codes together with their own optimal decoding metrics, and finally, compare to the trade-off function of fixed–rate codes.

Another objective is to examine the effect of imposing a privacy leakage constraint in addition to the FA error exponent constraint. This will be carried out in both the fixed–rate regime and the variable–rate regime.

IV. Optimal Rate Functions and Decoding Metrics

We begin by deriving optimal rate functions for both fixed–rate codes and variable–rate codes.

A. Fixed–Rate Codes

For fixed–rate codes, the following lemma establishes the optimal helper–message rate, $R_{\rm w}$, and secret key rate, $R_{\rm s}$, for a given value, $E_0 > 0$, of the FA error exponent, $E_{\rm FA}$.

Lemma 1 *Necessary and sufficient conditions for the existence of fixed–rate codes that achieve* $E_{FA}(R_w, R_s) \geq E_0$ *are:*

$$R_s \;\; \geq \;\; E_0, \tag{13}$$

$$
\begin{aligned}
R_{w} &\leq R_{w}^{*}(E_0) \\
&\triangleq \min_{\{Q_X:\ D(Q_X\|P_X)\leq E_0\}} \boldsymbol{E}_Q \log\frac{1}{P_X(X)} - E_0 \\
&= \sup_{\lambda\geq 0}\left\{-\lambda\ln\left(\sum_{x\in\mathcal{X}}[P_X(x)]^{1+1/\lambda}\right) - (1+\lambda)E_0\right\}. \qquad (14)
\end{aligned}
$$

Note that the requirement $R_{\rm s} \geq E_0$ is quite intuitive, because even a blind guess of $\boldsymbol{S}$ may succeed with probability of $e^{-nR_{\rm s}}$. It was shown in [13] that the best achievable FA exponent is given in turn by $I(X;Y)$. This is coherent with the result [6, Theorem 2.1] that $I(X;Y)$ is also an achievable upper bound on $R_{\rm s}$.

Proof of Lemma 1. From eq. (9), it is immediately seen that the statement, $E_{\rm FA}(R_{\rm w}, R_{\rm s}) \geq E_0$, is equivalent to the statement

$$
\forall\ Q_X \quad D(Q_X\|P_X) + \min\{R_{\rm s}, [H_Q(X) - R_{\rm w}]_+\} \geq E_0, \qquad (15)
$$

which in turn is equivalent to the two simultaneous statements,

$$
\forall\ Q_X \quad R_{\rm s} \geq E_0 - D(Q_X\|P_X) \qquad (16)
$$

$$
\forall\ Q_X \quad [H_Q(X) - R_{\rm w}]_+ \geq E_0 - D(Q_X\|P_X) \qquad (17)
$$

The former happens if and only if $R_{\rm s} \geq E_0$, which is eq. (13). As for the latter, for $D(Q_X\|P_X) \geq E_0$, the r.h.s. is non–positive, whereas the l.h.s. is non–negative, and so, there is no limitation on $R_{\rm w}$, which is associated with the region $\{Q_X:\ D(Q_X\|P_X) \geq E_0\}$. For $D(Q_X\|P_X) < E_0$, on the other hand, we must have $H_Q(X) - R_{\rm w} \geq E_0 - D(Q_X\|P_X)$, or equivalently,

$$
R_{\rm w} \leq H_Q(X) + D(Q_X\|P_X) - E_0 \equiv \boldsymbol{E}_Q \ln\frac{1}{P_X(X)} - E_0, \qquad (18)
$$

for every Q_X such that $D(Q_X\|P_X) < E_0$. This, in turn, is equivalent to the requirement given in the first two lines of eq. (14). The third line of (14) is obtained as follows:

$$
\begin{aligned}
R_{\rm w}^{*}(E_0) &= \min_{\{Q_X:\ D(Q_X\|P_X)\leq E_0\}} \boldsymbol{E}_Q \log\frac{1}{P_X(X)} - E_0 \\
&= \min_{Q_X}\sup_{\lambda\geq 0}\left\{\boldsymbol{E}_Q \log\frac{1}{P_X(X)} + \lambda[D(Q_X\|P_X) - E_0] - E_0\right\} \\
&= \sup_{\lambda\geq 0}\min_{Q_X}\left\{\boldsymbol{E}_Q \log\frac{1}{P_X(X)} + \lambda[D(Q_X\|P_X) - E_0] - E_0\right\}
\end{aligned}
$$

$$= \sup_{\lambda \geq 0} \left\{ -\lambda \ln \left(\sum_{x \in \mathcal{X}} [P_X(x)]^{1+1/\lambda} \right) - (1+\lambda) E_0 \right\}, \tag{19}$$

where the third equality is follows from convexity in Q_X and concavity (in fact, affinity) in λ.

B. Variable–Rate Codes

For variable–rate codes, we have the following lemma, which sets the stage for optimal rate functions.

Lemma 2 *Necessary and sufficient conditions for the existence of variable–rate codes that achieve FA error exponent at least as large as E_0 are:*

$$R_s(Q_X) \geq R_s^*(Q_X, E_0) \stackrel{\Delta}{=} E_0 - D(Q_X \| P_X), \tag{20}$$

$$\begin{aligned} R_w(Q_X) &\leq R_w^*(Q_X, E_0) \\ &\stackrel{\Delta}{=} \begin{cases} \boldsymbol{E}_Q \log \frac{1}{P_X(X)} - E_0 & D(Q_X \| P_X) < E_0 \\ \infty & D(Q_X \| P_X) \geq E_0. \end{cases} \end{aligned} \tag{21}$$

Observe that $R_{\rm s}^*(Q_X, E_0) + R_{\rm w}^*(Q_X, E_0) = H_Q(X)$ for all Q_X with $D(Q_X \| P_X) < E_0$, which roughly speaking, means that the mapping from $\boldsymbol{x}$ to $(\boldsymbol{s}, \boldsymbol{w})$ is one–to–one within each type, $\mathcal{T}(Q_X)$.

Proof of Lemma 2. Eq. (9) easily extends to the variable–rate case, by simply substituting $R_{\rm s}(Q_X)$ and $R_{\rm w}(Q_X)$ instead of $R_{\rm s}$ and $R_{\rm w}$, respectively. Therefore, the same reasoning as in the proof of Lemma 1 applies in the variable–rate setting considered here as well, except that now, there is no need for optimization (maximization, in the case of $R_{\rm s}$, and minimization, in the case of $R_{\rm w}$), as the binning rates are allowed to depend on the type, Q_X.

V. FR–FA Trade-off Functions

In this section, we characterize the FR–FA trade-off functions for both the random coding ensembles of both fixed–rate and variable–rate codes.

A. Fixed–Rate Codes

According to eq. (10), the random coding FR exponent depends on $R_{\rm w}$) only, and it is a monotonically, non–decreasing function of this variable. Thus, the best one can do with fixed–rate codes

is to use the highest allowable binning rate, which is $R_{\rm w}^*$. Therefore, if we denote the fixed–rate FR–FA trade-off function by $E_{\rm FR}^{\rm f}[E_0]$ (where the superscript "f" stands for "fixed–rate"), we have the following expression for the optimal decoding metric, $a(Q) = -H_Q(X|Y)$ (see eq. (12)),

$$E_{\rm FR}^{\rm f}[E_0] = E_{\rm FR}(R_{\rm w}^*(E_0)) = \min_{Q_{XY}} \{D(Q_{XY}\|P_{XY}) + [R_{\rm w}^*(E_0) - H_Q(X|Y)]_+\}, \tag{22}$$

which is also well known to be the Csiszár–style formula for the error exponent associated with ordinary MAP decoding (of the full source vector $\boldsymbol{X}$, rather than just the secret key $\boldsymbol{S}$) for a random Slepian–Wolf code (see, e.g., [3, eqs. (7), (19)]).

As mentioned already in the Introduction, in this paper, we are also be interested in Gallager–style forms, since they are more convenient to work with when it comes to numerical calculations. The Gallager–style form of eq. (22) is well known [5], [3, p. 9] to be

$$E_{\rm FR}^{\rm f}[E_0] = \max_{0\le\rho\le 1} \left\{ -\ln \left[\sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} [P_{XY}(x,y)]^{1/(1+\rho)} \right)^{1+\rho} \right] + \rho R_{\rm w}^*(E_0) \right\}. \tag{23}$$

Upon substituting the expression of $R_{\rm w}^*(E_0)$ (see eq. (14)), we finally obtain

$$\begin{aligned} E_{\rm FR}^{\rm f}[E_0] \;\; = \;\; & \max_{0\le\rho\le 1} \sup_{\lambda\ge 0} \left\{ -\ln \left[\sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} [P_{XY}(x,y)]^{1/(1+\rho)} \right)^{1+\rho} \right] - \right. \\ & \left. \rho\lambda \ln \left(\sum_{x\in\mathcal{X}} [P_X(x)]^{1+1/\lambda} \right) - \rho(1+\lambda)E_0 \right\}. \end{aligned} \tag{24}$$

B. Variable–Rate Codes

We now derive the FR–FA trade-off function for the variable–rate case, which will be denoted by $E_{\rm FR}^{\rm v}[E_0]$. Once again, since the FR exponent is monotonically non–decreasing in the helper–message rate, the best we can do is let $R_{\rm w}(Q_X) = R_{\rm w}^*(Q_X, E_0)$, as defined in Lemma 2.

The following point, however, should be considered carefully: the universal decoding metric $a(Q_{XY}) = -H_Q(X|Y)$, that we have used above for fixed–rate codes, is no longer equivalent to that of MAP decoding (and hence no longer optimal) for variable–rate codes. For a given rate function, $R_{\rm w}(Q_X)$, the following decoding metric should be used instead in order to obtain the same random coding FR exponent as in MAP decoding:

$$a(Q_{XY}) = R_{\rm w}(Q_X) - H_Q(X|Y). \tag{25}$$

In this case, referring to eqs. (10) and (12) of [9], we have

$$\begin{aligned} E(R_{\rm w}, Q_{X_0Y}) &= \inf_{Q_{X|Y}} [R_{\rm w}(Q_X) - H_Q(X|Y) + [a(Q_{X_0Y}) - a(Q_{XY})]_+]_+ \\ &= \inf_{Q_{X|Y}} [R_{\rm w}(Q_X) - H_Q(X|Y) + [R_{\rm w}(Q_{X_0}) - H_Q(X_0|Y) - \\ &\quad \{R_{\rm w}(Q_X) - H_Q(X|Y)\}]_+]_+ \\ &= \inf_{Q_{X|Y}} [\max\{R_{\rm w}(Q_X) - H_Q(X|Y), R_{\rm w}(Q_{X_0}) - H_Q(X_0|Y)\}]_+]_+ \\ &= [R_{\rm w}(Q_{X_0}) - H_Q(X_0|Y)]_+ \\ &\geq \min\{[R_{\rm w}(Q_X) - H_Q(X|Y)]_+ : \ \boldsymbol{E}_Q \ln P(X|Y) \geq \boldsymbol{E}_Q \ln P(X_0|Y)\}, \end{aligned} \tag{26}$$

where the last line corresponds to (pairwise) errors pertaining to the MAP decoder. Now, for $R_{\rm w}(Q_X) = R_{\rm w}^*(Q_X, E_0)$, we can present the FR error exponent (omitting the subscript 0 of X_0, which is no longer needed):

$$E_{\rm FR}^{\rm v}[E_0] = \min_{\{Q_{XY}:\ D(Q_X\|P_X)\leq E_0\}} \left\{ D(Q_{XY}\|P_{XY}) + \left[\boldsymbol{E}_Q \ln \frac{1}{P_X(X)} - E_0 - H_Q(X|Y) \right]_+ \right\}. \tag{27}$$

This is the Csiszár–style formula of the FR–FA trade-off function for variable–rate codes. The following theorem provides the Gallager–style form of the same function.

Theorem 1 *The variable–rate FR–FA trade-off function (27) can also be presented as*

$$\begin{aligned} E_{FR}^{v}[E_0] &= \max_{0\leq\lambda\leq 1} \sup_{\rho\geq 0} \max_{W} \left\{ -\ln \left(\sum_{y\in\mathcal{W}} \left[\sum_{x\in\mathcal{X}} [P_{XY}(x,y) P_X(x)^{\rho+\lambda} W(x)^{-\rho}]^{1/(1+\lambda)} \right]^{1+\lambda} \right) - \right. \\ &\quad \left. (\rho+\lambda) E_0 \right\}, \end{aligned} \tag{28}$$

where the maximum over W is taken over the simplex of probability distributions over $\mathcal{X}$, i.e., $W(x) \geq 0$ for all $x \in \mathcal{X}$ and $\sum_{x\in\mathcal{X}} W(x) = 1$.

Before turning to the proof of Theorem 1, we pause to demonstrate it and to discuss some aspects, consequences and extensions of this theorem.

Optimization issues. First, note that the formula (28) involves optimization over a probability distribution W in addition to the parameters ρ and λ, namely, a total of $|\mathcal{X}| + 1$ parameters. This number is never larger (and in most cases, considerably smaller) than the $|\mathcal{X}|\cdot|\mathcal{Y}|-1$ parameters that are associated with the minimization over Q_{XY} in the Csiszár–style formula of eq. (27). Moreover,

since the Gallager–style formula (28) involves only maximization, any arbitrary choice of λ, ρ and W in their allowed ranges, would yield a valid lower bound (a guarantee) on the achievable FR error exponent. This is different from the situation with the Csiszár–style formula, which involves minimization, and hence allows no such privilege: one **must** carry out the minimization in order to obtain the achievable FR error exponent. One drawback of the Gallager–style formula is that the range of maximization over the parameter ρ is infinite. In practical numerical calculations, however, one can initially limit the range to an interval of the form $[0, \rho_0]$, and then gradually enlarge ρ_0 up to the point where no further increase in ρ_0 improves on the resulting maximum.

A few words are in order concerning the maximization over W, which is a relatively computationally demanding step, especially for a large source alphabet. First, note that in situations with a sufficient degree of symmetry, the optimal W turns out to be the uniform distribution, a fact that saves the optimization numerically. This turns out to be the case if P_X is uniform and the $|\mathcal{Y}|$ probability vectors $\{P_{X|Y=y}(x|y),\ x \in \mathcal{X}\}$, are all permutations, $\{\pi_y\}$, of one such vector, that form a group (w.r.t. compositions of permutations), such that $(1/|\mathcal{Y}|)\sum_y \pi_y^{-1}[W] = U$, where U designates the uniform distribution over $\mathcal{X}$. This happens, for instance, when $P_X = U$ and $P_{X|Y}$ is a modulo–additive channel. To see why this is true, we define

$$\begin{aligned} f(W) \quad = \quad & \inf_{Q_{XY}} \sum_{y\in\mathcal{Y}} Q_Y(y) \ln \frac{Q_Y(y)}{P_Y(y)} + \sum_{y\in\mathcal{Y}} Q_Y(y) \sum_x Q_{X|Y}(x|y) \left[\ln \frac{Q_{X|Y}(x|y)}{P_{X|Y}(x|y)} + \right. \\ & \left. (\rho+\lambda) \ln \frac{1}{P_X(x)} + \lambda \ln Q_{X|Y}(x|y) + \rho \ln W(x) \right], \end{aligned} \tag{29}$$

which we show[1] to be equal to

$$f(W) = -\ln \left(\sum_{y\in\mathcal{Y}} \left[\sum_{x\in\mathcal{X}} [P_{XY}(x,y) P_X(x)^{\rho+\lambda} W(x)^{-\rho}]^{1/(1+\lambda)} \right]^{1+\lambda} \right). \tag{30}$$

From the first representation of f, it is easy to see that it is concave in W. From the second representation and the assumed symmetry, it is easy to see that for every W and every $y \in \mathcal{Y}$, $f(\pi_y[W]) = f(W)$. Thus,

$$f(W) = \frac{1}{|\mathcal{Y}|} \sum_{y\in\mathcal{Y}} f(\pi_y[W]) \le f\left(\frac{1}{|\mathcal{Y}|} \sum_{y\in\mathcal{Y}} \pi_y[W] \right) = f(U), \tag{31}$$

[1] See the proof of Theorem 1 in the sequel.

which means that the optimal W is uniform. In the general case, the optimization over W is a convex program, and so, there are standard solvers that can handle this problem more efficiently than a brute–force exhaustive search.

Example. In order to demonstrate the advantage of variable–rate codes relative to fixed–rate codes in terms of the FR–FA trade-off, we now provide a simple example. Consider the case of a double binary source with alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and joint probabilities given by $P_{XY}(0, 0) = 0.32$, $P_{XY}(0, 1) = 0.08$, $P_{XY}(1, 0) = 0.06$, and $P_{XY}(1, 1) = 0.54$. Fig. 1 displays the two FR–FA trade-off functions, $E^{\mathrm{f}}_{\mathrm{FR}}[E_0]$ and $E^{\mathrm{v}}_{\mathrm{FR}}[E_0]$, for this source. As can be seen, the gap between these two functions is rather considerable, which means that variable–rate codes with optimal rate functions, are significantly better in terms of these trade–offs. This example is quite representative in the sense that other examples (with different source probabilities) yielded qualitatively similar results.

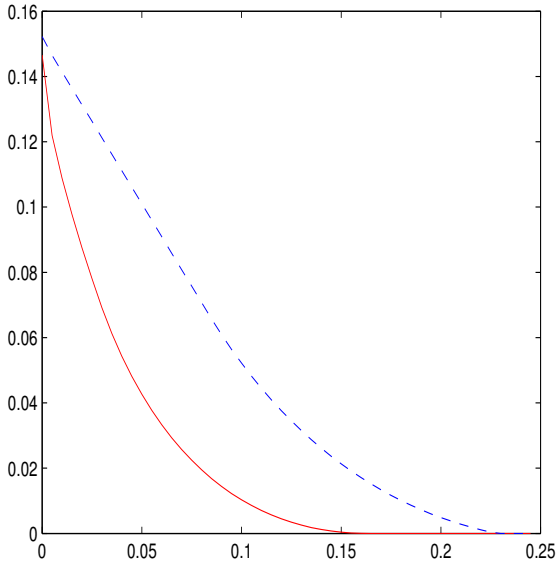

Figure 1: Graphs of $E^{\mathrm{f}}_{\mathrm{FR}}[E_0]$ (solid red curve) and $E^{\mathrm{v}}_{\mathrm{FR}}[E_0]$ (dashed blue curve) for the double binary source, defined by $P_{XY}(0, 0) = 0.32$, $P_{XY}(0, 1) = 0.08$, $P_{XY}(1, 0) = 0.06$, and $P_{XY}(1, 1) = 0.54$.

Variable–rate codes do not always improve on fixed–rate codes. It should be kept in mind, however, that there are situations where variable–rate codes offer no improvement over fixed–rate codes, i.e., they might have exactly the same FR–FA trade-off function in some cases.

One such example is the case where the source X has a uniform distribution. In this case, the optimal rate function for variable–rate codes turns out to be $R_{\rm w}^*(Q_X, E_0) = \ln|\mathcal{X}| - E_0$ (whenever $D(Q_X\|P_X) < E_0$), which is independent of Q_X, and hence is a fixed–rate anyway. So unless the dominant type Q_X happens to fall in the region $\{Q_X : \ D(Q_X\|P_X) \geq E_0\}$, when the source is uniform, variable–rate codes cannot offer any improvement beyond the performance of fixed–rate codes.

Another aspect is associated with the decoding metric. Even for a general source, P_{XY}, if one uses a random variable–rate code, but decodes it using the decoding metric function of fixed–rate codes, $a(Q_{XY}) = -H_Q(X|Y)$, instead of the optimal decoding metric for variable–rate codes, $a(Q_{XY}) = R_{\rm w}^*(Q_X, E_0) - H_Q(X|Y)$, then the resulting FR–FA trade–off function turns out to be exactly the same as with fixed–rate codes.

Mismatched decoding. Our results can be extended to apply to a mismatched decoding metric, $a(Q_{XY}) = \boldsymbol{E}_Q \ln P'(X|Y)$ (see eq. (6)), for an arbitrary P', using the same techniques. The resulting fixed–rate Gallager–style FR–FA trade-off function would then be

$$E_{\rm FR}^{\rm f}[E_0] = \max_{0\leq s\leq 1}\max_{0\leq t\leq s}\left\{-\ln\left(\sum_{y\in\mathcal{Y}}\left[\left(\sum_{x\in\mathcal{X}}[P'(x|y)]^{t/s}\right)^s\cdot\left(\sum_{x'\in\mathcal{X}}\frac{P_{XY}(x',y)}{[P'(x'|y)]^t}\right)\right]\right)+ sR_{\rm w}^*(E_0)\right\}, \tag{32}$$

where $R_{\rm w}^*(E_0)$ is as defined in Lemma 1. The corresponding variable–rate function, which cannot be smaller, is given by

$$E_{\rm FR}^{\rm v}[E_0] = \sup_{\lambda\geq 0}\max_{0\leq s\leq 1}\max_{0\leq t\leq s}\max_{W}\left\{-\ln\left(\sum_{y\in\mathcal{Y}}\left[\left(\sum_{x\in\mathcal{X}}[P'(x|y)]^{t/s}[P_X(x)]^{1+\lambda/s}[W(x)]^{-\lambda/s}\right)^s\times\left(\sum_{x'\in\mathcal{X}}\frac{P_{XY}(x',y)}{[P'(x'|y)]^t}\right)\right]\right)-(\lambda+s)E_0\right\}. \tag{33}$$

Expurgated exponents. In [9], expurgated FR exponents were also derived, and so, in principle, one could carry out similar analyses for trade–offs between expurgated exponents and the best achievable FA error exponents. We have not pursued such derivations in this work since they are significantly more complicated, but it is anticipated that similar conclusions would apply concerning the advantage of variable–rate codes over fixed–rate codes. In this context, it should be emphasized

that one of our important messages in this work is not only that variable–rate codes are better than fixed–rate codes, but that moreover, we characterize the *optimal rate functions*, independently of the type of FR error exponents being considered (random coding exponents, expurgated exponents, sphere–packing exponent [9, Theorem 3] at a certain rate region, etc.) since their derivation stems from the FA error exponent, which has an exact characterization [9, Section V].

The remaining part of this section is devoted to the proof of Theorem 1.

Proof of Theorem 1. Throughout this proof we will make frequent use of the minimax theorem, based on convexity–concavity arguments. We will also use repeatedly the fact that

$$\min_{Q}[D(Q\|P)+\boldsymbol{E}_Q f(X)] = -\ln\left[\sum_x P(x)e^{-f(x)}\right].$$

Now,

$$\begin{aligned}
E_{\rm FR}^{\rm v}[E_0] &= \inf_{Q_{XY}}\sup_{0\le\lambda\le 1}\sup_{\rho\ge 0}\Bigg\{D(Q_{XY}\|P_{XY})+\lambda\left[\boldsymbol{E}_Q\ln\frac{1}{P_X(X)}-E_0-H_Q(X|Y)\right]+\cdot\\
&\quad \rho\left[D(Q_X\|P_X)-E_0\right]\Bigg\}\\
&= \sup_{0\le\lambda\le 1}\sup_{\rho\ge 0}\inf_{Q_{XY}}\Bigg\{D(Q_{XY}\|P_{XY})+\lambda\left[\boldsymbol{E}_Q\ln\frac{1}{P_X(X)}-E_0-H_Q(X|Y)\right]+\\
&\quad \rho\left[D(Q_X\|P_X)-E_0\right]\Bigg\}\\
&= \sup_{0\le\lambda\le 1}\sup_{\rho\ge 0}\inf_{Q_Y}\Bigg\{D(Q_Y\|P_Y)+\inf_{Q_{X|Y}}\Big(D(Q_{X|Y}\|P_{X|Y}|Q_Y)+\\
&\quad \lambda\left[\boldsymbol{E}_Q\ln\frac{1}{P_X(X)}-E_0-H_Q(X|Y)\right]+\rho\left[D(Q_X\|P_X)-E_0\right]\Big)\Bigg\}\\
&= \sup_{0\le\lambda\le 1}\sup_{\rho\ge 0}\inf_{Q_Y}\Bigg\{D(Q_Y\|P_Y)+\inf_{Q_{X|Y}}\sum_{y\in\mathcal{Y}}Q_Y(y)\sum_x Q_{X|Y}(x|y)\left[\ln\frac{Q_{X|Y}(x|y)}{P_{X|Y}(x|y)}+\right.\\
&\quad \left.(\rho+\lambda)\ln\frac{1}{P_X(x)}+\lambda\ln Q_{X|Y}(x|y)+\rho\ln Q_X(x)\right]-(\rho+\lambda)E_0\Bigg\}. \qquad (34)
\end{aligned}$$

Consider first the inner–most minimization over $\{Q_{X|Y}\}$:

$$\begin{aligned}
&\inf_{Q_{X|Y}}\sum_{y\in\mathcal{Y}}Q_Y(y)\sum_{x\in\mathcal{X}}Q_{X|Y}(x|y)\left[\ln\frac{Q_{X|Y}(x|y)}{P_{X|Y}(x|y)}+(\rho+\lambda)\ln\frac{1}{P_X(x)}+\lambda\ln Q_{X|Y}(x|y)+\rho\ln Q_X(x)\right]\\
&= \inf_{Q_{X|Y}}\sup_{W}\sum_{y\in\mathcal{Y}}Q_Y(y)\sum_{x\in\mathcal{X}}Q_{X|Y}(x|y)\left[\ln\frac{Q_{X|Y}(x|y)}{P_{X|Y}(x|y)}+(\rho+\lambda)\ln\frac{1}{P_X(x)}+\lambda\ln Q_{X|Y}(x|y)+\rho\ln W(x)\right]\\
&= \sup_{W}\sum_{y\in\mathcal{Y}}Q_Y(y)\inf_{Q_{X|Y}}\sum_{x\in\mathcal{X}}Q_{X|Y}(x|y)\left[\ln\frac{Q_{X|Y}(x|y)}{P_{X|Y}(x|y)}+(\rho+\lambda)\ln\frac{1}{P_X(x)}+\lambda\ln Q_{X|Y}(x|y)+\rho\ln W(x)\right]
\end{aligned}$$

$$= (1+\lambda)\sup_W \sum_{y\in\mathcal{Y}} Q_Y(y) \inf_{Q_{X|Y}} \sum_{x\in\mathcal{X}} Q_{X|Y}(x|y) \ln \frac{Q_{X|Y}(x|y)}{[P_{X|Y}(x|y)P_X(x)^{\rho+\lambda}W(x)^{-\rho}]^{1/(1+\lambda)}}$$

$$= -(1+\lambda)\inf_W \sum_y Q(y) \ln\left(\sum_x [P(x|y)P(x)^{\rho+\lambda}W(x)^{-\rho}]^{1/(1+\lambda)}\right), \tag{35}$$

which, after the minimization over Q_Y, becomes

$$\inf_{Q_Y}\left[D(Q_Y\|P_Y) - (1+\lambda)\inf_W \sum_{y\in\mathcal{Y}} Q_Y(y)\ln\left(\sum_{x\in\mathcal{X}}[P_{X|Y}(x|y)P_X(x)^{\rho+\lambda}W(x)^{-\rho}]^{1/(1+\lambda)}\right)\right]$$

$$= -\inf_W \ln\left(\sum_{y\in\mathcal{Y}}\left[\sum_{x\in\mathcal{X}}[P_{XY}(x,y)P_X(x)^{\rho+\lambda}W(x)^{-\rho}]^{1/(1+\lambda)}\right]^{1+\lambda}\right). \tag{36}$$

It follows that

$$E_{\rm FR}^{\rm v}[E_0] = \max_{0\le\lambda\le 1}\sup_{\rho\ge 0}\sup_W\left\{-\ln\left(\sum_{y\in\mathcal{Y}}\left[\sum_{x\in\mathcal{X}}[P_{XY}(x,y)P_X(x)^{\rho+\lambda}W(x)^{-\rho}]^{1/(1+\lambda)}\right]^{1+\lambda}\right) - (\rho+\lambda)E_0\right\}.$$

This completes the proof of Theorem 1.□

VI. Privacy Leakage

The privacy leakage of the system is defined as $I(\boldsymbol{X};\boldsymbol{W})$. Since $\boldsymbol{W}$ is a deterministic function of $\boldsymbol{X}$, we have $I(\boldsymbol{X};\boldsymbol{W}) = H(\boldsymbol{W}) = -\boldsymbol{E}\ln P(\boldsymbol{W})$.

Let $N(Q_X,\boldsymbol{w})$ be the number of $\boldsymbol{x}$–vectors of type Q_X for which $f(\boldsymbol{x})=\boldsymbol{w}$. Clearly, $N(Q_X,\boldsymbol{w})$ is a binomial random variable with $|\mathcal{T}(Q_X)| \doteq e^{nH_Q(X)}$ trials and success rate of $e^{-nR_{\rm w}}$. Now, for the typical code,

$$\begin{aligned} P(\boldsymbol{w}) &= \sum_{\boldsymbol{x}} P(\boldsymbol{x})\mathcal{I}\{f(\boldsymbol{x})=\boldsymbol{w}\} \\ &= \sum_{Q_X} P(\boldsymbol{x})\Big|_{\boldsymbol{x}\in\mathcal{T}(Q_X)} \cdot N(Q_X,\boldsymbol{w}) \\ &\doteq \sum_{\{Q_X:\ H(Q_X)\ge R_{\rm w}(Q_X)\}} e^{-n[H(Q_X)+D(Q_X\|P_X)]}\cdot e^{n[H(Q_X)-R_{\rm w}(Q_X)]} \\ &\doteq \exp\left\{-n\min_{\{Q_X:\ H(Q_X)\ge R_w(Q_X)\}}[R_{\rm w}(Q_X)+D(Q_X\|P_X)]\right\}, \end{aligned} \tag{37}$$

independently of $\boldsymbol{w}$, and so, for the typical code

$$H(\boldsymbol{W}) \approx \min_{\{Q_X:\ H(Q_X)\ge R_{\rm w}(Q_X)\}}[R_{\rm w}(Q_X)+D(Q_X\|P_X)]. \tag{38}$$

If we impose the constraint $H(\boldsymbol{W}) \leq nH_0$ (in addition to the FA constraint), for some prescribed constant H_0, then this constraint is equivalent to the requirement that there would exist at least one pmf Q_X for which $R_{\rm w}(Q_X) \leq \min\{H(Q_X), H_0 - D(Q_X\|P_X)\}$. In the variable–rate case, as long as $H_0 \geq E_0$ and $E_0 \leq -\ln\min_x P_X(x)$, there is no tension between this requirement and the requirement that comes from FA exponent constraint since we can pick $R_{\rm w}(Q_X)$ as we did before and comply with the new requirement for some Q_X with $D(Q_X\|P_X) > E_0$, which is the region where the FA constraint poses no limitations. This is, of course, possible only as long as $H_0 > E_0$ (otherwise the upper bound on $R_{\rm w}(Q_X)$ is negative in the entire region $\{Q_X : D(Q_X\|P_X) > E_0\}$) and $E_0 \leq -\ln\min_x P_X(x)$ (otherwise that region is empty). Recall that E_0 cannot exceed $R_{\rm s}$, which in turn cannot exceed $I(X;Y)$.

Alternatively, for the privacy leakage constraint to be inactive, in the context of the previous analysis for variable rate codes, it is enough that there would be at least one Q_X with $D(Q_X\|P_X) \leq E_0$ such that

$$\boldsymbol{E}_Q \ln\frac{1}{P_X(X)} - E_0 \leq H_0 - D(Q_X||P_X). \tag{39}$$

Since H_0 must be at least as large as $H(X|Y)$ (see the discussion in the Introduction, as well as [6, Proposition 2.4]) and E_0 can be made close to $I(X;Y)$, $H_0 + E_0$ can be close to $H(X)$, in which case, $Q_X = P_X$ is a feasible choice. Otherwise, the condition for the existence of such Q_X is

$$\min_{\{Q_X:\ D(Q_X\|P_X)\leq E_0\}}\left[\boldsymbol{E}_Q \ln\frac{1}{P_X(X)} + D(Q_X\|P_X)\right] \leq H_0 + E_0, \tag{40}$$

or, equivalently,

$$\sup_{\lambda\geq 1}\left\{-\lambda\ln\left[\sum_x P_X^{1+1/\lambda}(x)\right] - \lambda E_0\right\} \leq H_0. \tag{41}$$

Thus, as long as H_0 is not too small (depending on E_0), the privacy leakage constraint does not affect the earlier analysis.

In the fixed–rate case, on the other hand, the situation is somewhat different. Here the privacy leakage constraint requires

$$\begin{aligned} R_{\rm w} &\leq \max_{Q_X}\min\{H_Q(X), H_0 - D(Q_X\|P_X)\} \\ &= \max_{Q_X}\min_{0\leq s\leq 1}\{(1-s)H_Q(X) + s[H_0 - D(Q_X\|P_X)]\} \\ &= \min_{0\leq s\leq 1}\max_{Q_X}\{(1-s)H_Q(X) + s[H_0 - D(Q_X\|P_X)]\} \end{aligned}$$

$$
\begin{aligned}
&= \min_{0\le s\le 1} \max_{Q} \sum_{x\in\mathcal{X}} Q_X(x) \left\{ -(1-s)\ln Q_X(x) + s\ln\frac{P_X(x)}{Q_X(x)} + sH_0 \right\} \\
&= \min_{0\le s\le 1} \max_{Q_X} \sum_{x\in\mathcal{X}} Q_X(x) \left\{ \ln\frac{P_X^s(x)}{Q_X(x)} + sH_0 \right\} \\
&= \min_{0\le s\le 1} \left\{ \ln\left[\sum_{x\in\mathcal{X}} P_X^s(x)\right] + sH_0 \right\} \\
&\stackrel{\Delta}{=} R_{\mathrm{w}}^*(H_0). \qquad (42)
\end{aligned}
$$

Of course, ultimately R_{w} is not allowed to exceed $\min\{R_{\mathrm{w}}^*(E_0), R_{\mathrm{w}}^*(H_0)\}$. Thus, when the privacy leakage constraint is added into the picture, this increases the gap between the variable–rate and the fixed–rate regimes even further.

References

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography – part I: secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, July 1993.

[2] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography – part II: CR capacity," *IEEE Trans. Inform. Theory*, vol. 44, pp. 225–240, January 1998.

[3] J. Chen, D.–k He, A. Jagmohan, and A. Lastras-Montaño, "On the reliability of variable–rate Slepian–Wolf coding," *Entropy*, vol. 19, 389, 2017. doi:10.3390/e19080389

[4] I. Csiszár and J. Körner, "Graph decomposition: a new key to coding theorems," *IEEE Trans. Inform. Theory*, vol. IT–27, no. 1, pp. 5–12, January 1981.

[5] R. G. Gallager, "Source coding with side information and universal coding," LIDS-P-937, M.I.T., 1976.

[6] T. Ignatenko and F. M. J. Willems, "Biometric security from an information–theoretical perspective," *Foundations and Trends in Communications and Information Theory*, vol. 7, nos. 2–3, 2010.

[7] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, May 1993.

[8] N. Merhav, "The generalized stochastic likelihood decoder: random coding and expurgated bounds," *IEEE Trans. Inform. Theory*, vol. 63, no. 8, pp. 5039–5051, August 2017.

[9] N. Merhav, "Ensemble performance of biometric authentication systems based on secret key generation," submitted to *IEEE Trans. Inform. Theory*, July 2017.
Available on–line at: `http://webee.technion.ac.il/people/merhav/papers/p199r.pdf`

[10] J. Scarlett, A. Martinéz and A. G. i Fábregas, "The likelihood decoder: error exponents and mismatch," *Proc. 2015 IEEE International Symposium on Information Theory (ISIT 2015)*, pp. 86–90, Hong Kong, June 2015.

[11] E. C. Song, P. Cuff and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inform. Theory*, vol. 62, no. 4, pp. 1836–1849, April 2016.

[12] *N. Weinberger* and N. Merhav, "Optimum tradeoffs between the error exponent and the excess–rate exponent of variable–rate Slepian–Wolf coding," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 2165–2190, April 2015.

[13] F. M. J. Willems and T. Ignatenko, "Authentication based on secret–key generation," *2012 IEEE Proc. International Symposium on Information Theory (ISIT 2012)*, pp. 1792–1796, Cambridge, MA, U.S.A, July 2012.

[14] M. H. Yassaee, M. R. Aref and A. Gohari, "A technique for deriving one–shot achievability results in network information theory," *Proc. 2013 IEEE International Symposium on Information Theory (ISIT 2013)*, pp. 1287–1291, July 2013. Also, available on–line at `http://arxiv.org/abs/1303.0696`.