Distributed Storage for Data Security

Annina Bracher
ETH Zurich

Eran Hof
Samsung Israel Research and Development Center

Amos Lapidoth
ETH Zurich

***Abstract*—We study the secrecy of a distributed storage system for passwords. The encoder, Alice, observes a length-n password and describes it using two hints, which she then stores in different locations. The legitimate receiver, Bob, observes both hints. In one scenario we require that the number of guesses it takes Bob to guess the password approach 1 as n tends to infinity and in the other that the size of the list that Bob must form to guarantee that it contain the password approach 1. The eavesdropper, Eve, sees only one of the hints; Alice cannot control which. For each scenario we characterize the largest normalized (by n) exponent that we can guarantee for the number of guesses it takes Eve to guess the password.**

I. Introduction

Suppose that some sensitive information X (e.g. password) is drawn from a finite set $\mathcal{X}$ according to some PMF P_X. A (stochastic) encoder, Alice, maps (possibly using randomization) X to two hints M_1 and M_2, which she then stores in different locations. The hints are intended for a legitimate receiver, Bob, who knows where they are stored and has access to both. An eavesdropper, Eve, sees one of the hints but not both; we do not know which. Given some notion of ambiguity, we would ideally like Bob's ambiguity about X to be small and Eve's large.

Which hint is revealed to Eve is a subtle question. We adopt a conservative approach and assume that, after observing X, an adversarial genie reveals to Eve the hint that minimizes her ambiguity. Not allowing the genie to observe X would lead to a weaker form of secrecy (an example is given in [1]).

There are several ways to define ambiguity. For example, we could require that Bob be able to reconstruct X whenever X is "typical" and that the conditional entropy of X given Eve's observation be large. For some scenarios, such an approach might be inadequate. Firstly, this approach may not properly address Bob's needs when X is not typical. For example, if Bob must guess X, this approach does not guarantee that the expected number of guesses be small: It only guarantees that the probability of success after one guess be large. It does not indicate the number of guesses that Bob might need when X is atypical. Secondly, conditional entropy need not be an adequate measure of Eve's ambiguity: if X is some password that Eve wishes to uncover, then we may care more about the number of guesses that Eve needs than about the conditional entropy [2].

In this paper, we assume that Eve wants to guess X with the least number of guesses of the form "Is X = x?". We quantify Eve's ambiguity about X by the expected number of guesses that she needs to uncover X. In this sense, Eve faces an instance of the Massey-Arikan guessing problem [3], [4]. For each possible observation z in some finite set $\mathcal{Z}$, Eve chooses a guessing function $G(\cdot\,|z)$ from $\mathcal{X}$ onto the set $\{1,\ldots,|\mathcal{X}|\}$, which determines the guessing order: if Eve observes z, then the question "Is $X = x$?" will be her $G(x|z)$-th question. Eve's expected number of guesses is $\mathbb{E}[G(X|Z)]$. This expectation is minimized if for each $z \in \mathcal{Z}$ the guessing function $G(\cdot\,|z)$ orders the possible realizations of X in decreasing order of their posterior probabilities given $Z = z$.

As to Bob, we will consider two different criteria: In the "guessing version" of the problem the criterion is the expected number of guesses it takes Bob to guess X, and in the "list version" the criterion is the first moment of the size of the list that Bob must form to guarantee that it contain X.[1] We shall see that the two criteria lead to similar results.

The former criterion is natural when Bob can check whether a guess is correct: If X is some password, then Bob can stop guessing as soon as he has gained access to the account that is secured by X.

The latter criterion is appropriate if Bob does not know whether a guess is correct. For example, if X is a task that Bob must perform, then the only way for Bob to make sure that he performs X is to perform all the tasks in a list comprising the tasks that have positive posterior probabilities given his observation. In this scenario, a good measure of Bob's ambiguity about X is the expected number of tasks that he must perform, and this will be small whenever Alice is a good task-encoder for Bob [5]. To describe the list of tasks that Bob must perform more explicitly, let us denote by

$$\mathbb{P}[M_1 = m_1, M_2 = m_2\,|X = x]\,,\ m_1 \in \mathcal{M}_1,\ m_2 \in \mathcal{M}_2$$

the probability that Alice produces the pair of hints $(M_1, M_2) = (m_1, m_2)$ upon observing that $X = x$. It is 0-1 valued if Alice does not use randomization. Upon observing that $(M_1, M_2) = (m_1, m_2)$, Bob produces the list $\mathcal{L}_{m_1,m_2}$ of all the tasks $x \in \mathcal{X}$ whose posterior probability $\mathbb{P}[X = x|M_1 = m_1, M_2 = m_2]$ is positive. Our notion of Bob's ambiguity about X is $\mathbb{E}\big[|\mathcal{L}_{M_1,M_2}|\big]$.

The guessing and the list-size criterion for Bob lead to similar results in the following sense: Clearly, every guessing function $G(\cdot|M_1, M_2)$ for X that maps the elements of $\mathcal{X}$ that have zero posterior probability to larger values

[1] Our setup differs from the one in [2] in the following sense: Instead of mapping X to a public message using a secret key, which is available to Bob but not to Eve, here Alice produces two hints and stores them so that Bob sees both but Eve sees only one. Moreover, unlike [2] we do not measure Bob's ambiguity in terms of the probability that X is not his first guess.

than those that have a positive posterior probability satisfies $\mathbb{E}\big[G(X|M_1,M_2)\big] \le \mathbb{E}\big[|\mathcal{L}_{M_1,M_2}|\big]$. Conversely, one can prove that every pair of ambiguities for Bob and Eve that is achievable in the "guessing version" is, up to polylogarithmic factors of $|\mathcal{X}|$, also achievable in the "list version" provided that we increase $\mathcal{M}_1$ or $\mathcal{M}_2$ by a logarithmic factor of $|\mathcal{X}|$ [1]. These polylogarithmic factors wash out in the asymptotic regime where the sensitive information is an n-tuple and n tends to infinity.

With no extra effort we can generalize the model and replace expectations with ρ-th moments. This we do to better bring out the role of Rényi entropy. For an arbitrary $\rho > 0$, we thus study the ρ-th (instead of the first) moment of the list-size and of the number of guesses. Moreover, we shall allow some side-information Y that is available to all parties. We shall thus assume that the pair (X,Y) takes value in the finite set $\mathcal{X}\times\mathcal{Y}$ according to $P_{X,Y}$.

II. Problem Statement

We consider two problems, which we call the "guessing version" and the "list version". They differ in the definition of Bob's ambiguity. In both versions a pair (X,Y) is drawn from the finite set $\mathcal{X}\times\mathcal{Y}$ according to the PMF $P_{X,Y}$, and $\rho > 0$ is fixed. Upon observing $(X,Y) = (x,y)$, Alice draws the hints M_1 and M_2 from the finite set $\mathcal{M}_1\times\mathcal{M}_2$ according to some conditional PMF

$$\mathbb{P}[M_1 = m_1, M_2 = m_2 \,|X = x, Y = y]\,. \tag{1}$$

In the "guessing version" Bob's ambiguity about X is

$$\mathscr{A}_{\mathrm{B}}^{(\mathrm{g})}(P_{X,Y}) = \min_{G} \mathbb{E}\big[G(X|Y,M_1,M_2)^\rho\big]. \tag{2}$$

In the "list version" Bob's ambiguity about X is

$$\mathscr{A}_{\mathrm{B}}^{(\mathrm{l})}(P_{X,Y}) = \mathbb{E}\big[|\mathcal{L}^Y_{M_1,M_2}|^\rho\big], \tag{3}$$

where for all $y\in\mathcal{Y}$ and $(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2$

$$\mathcal{L}^y_{m_1,m_2} = \{x\colon \mathbb{P}[X = x\,|Y = y, M_1 = m_1, M_2 = m_2] > 0\}$$

is the list of all the tasks whose posterior probability

$$\begin{aligned}&\mathbb{P}[X = x\,|Y = y, M_1 = m_1, M_2 = m_2\,]\\ &= \frac{P_{X,Y}(x,y)\,\mathbb{P}[M_1{=}m_1, M_2{=}m_2\,|X{=}x, Y{=}y]}{\sum_{\tilde{x}} P_{X,Y}(\tilde{x},y)\,\mathbb{P}[M_1{=}m_1, M_2{=}m_2\,|X{=}\tilde{x}, Y{=}y]}\end{aligned} \tag{4}$$

is positive. In both versions Eve's ambiguity about X is

$$\mathscr{A}_{\mathrm{E}}(P_{X,Y}) = \min_{G_1,G_2} \mathbb{E}\big[G_1(X\,|Y,M_1\,)^\rho \wedge G_2(X\,|Y,M_2\,)^\rho\big], \tag{5}$$

where $\alpha\wedge\beta$ denotes the minimum of α and β.

Optimizing over Alice's mapping, i.e., the choice of the conditional PMF in (1), we wish to characterize the largest ambiguity that we can guarantee that Eve will have subject to a given upper bound on the ambiguity that Bob may have.

Of special interest to us is the asymptotic regime where (X,Y) is an n-tuple (not necessarily drawn IID), and where

$$\mathcal{M}_1 = \big\{1,\dots,2^{nR_1}\big\}\,,\ \mathcal{M}_2 = \big\{1,\dots,2^{nR_2}\big\}\,,$$

where (R_1,R_2) is a nonnegative pair corresponding to the rate. For both versions of the problem, we shall characterize the largest exponential growth that we can guarantee for Eve's ambiguity subject to the constraint that Bob's ambiguity tend to one. This asymptote turns out not to depend on the version of the problem, and in the asymptotic analysis $\mathscr{A}_{\mathrm{B}}$ can stand for either $\mathscr{A}_{\mathrm{B}}^{(\mathrm{g})}$ or $\mathscr{A}_{\mathrm{B}}^{(\mathrm{l})}$.

To phrase this mathematically, let us introduce the stochastic process $\{(X_i,Y_i)\}_{i\in\mathbb{N}}$ with finite alphabet $\mathcal{X}\times\mathcal{Y}$. We denote by P_{X^n,Y^n} the PMF of (X^n,Y^n). For a nonnegative rate-pair (R_1,R_2), we call E_{E} an *achievable ambiguity-exponent* if there is a sequence of stochastic encoders such that Bob's ambiguity (which is always at least 1) satisfies

$$\lim_{n\to\infty} \mathscr{A}_{\mathrm{B}}(P_{X^n,Y^n}) = 1, \tag{6}$$

and such that Eve's ambiguity satisfies

$$\liminf_{n\to\infty} \frac{\log(\mathscr{A}_{\mathrm{E}}(P_{X^n,Y^n}))}{n} \ge E_{\mathrm{E}}. \tag{7}$$

We shall characterize the supremum $\overline{E_{\mathrm{E}}}$ of all achievable ambiguity-exponents, which we call *privacy-exponent*. If (6) cannot be satisfied, then the set of achievable ambiguity-exponents is empty, and we say that the privacy-exponent is negative infinity.

III. Main Results

To describe our results, we shall need a conditional version of Rényi entropy (originally proposed by Arimoto [6] and also studied in [5])

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha}\log\sum_{y\in\mathcal{Y}}\Big(\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha\Big)^{1/\alpha}, \tag{8}$$

where $\alpha\in[0,\infty]$ is the order and where the cases where α is 0, 1, or ∞ are treated by a limiting argument. In addition, we shall need the notion of conditional Rényi entropy-rate: Let $\{(X_i,Y_i)\}_{i\in\mathbb{N}}$ be a discrete-time stochastic process with finite alphabet $\mathcal{X}\times\mathcal{Y}$. Whenever the limit as n tends to infinity of $H_\alpha(X^n|Y^n)\,/n$ exists, we denote it by $H_\alpha(\boldsymbol{X}|\boldsymbol{Y})$ and call it conditional Rényi entropy-rate. In this paper, $\alpha = 1/(1+\rho)$ takes value in the set $(0,1)$. To simplify notation, we henceforth write $\tilde{\rho}$ for $1/(1+\rho)$ and $\alpha\vee\beta$ for the maximum of α and β.

A. Finite Blocklength Results

In the next two theorems c_s is related to how much can be gleaned about X from (M_1,M_2) but not from one hint alone; c_1 is related to how much can be gleaned from M_1; and c_2 is related to how much can be gleaned from M_2. More precisely, we shall see in Section V ahead that Alice first maps (X,Y) to the triple (V_s,V_1,V_2), which takes value in a finite set $\mathcal{V}_s\times\mathcal{V}_1\times\mathcal{V}_2$, where $|\mathcal{V}_\nu| = c_\nu$, $\nu\in\{s,1,2\}$. Independently of (X,Y) she then draws a (one-time-pad like) random variable U uniformly over $\mathcal{V}_s$ and maps (U,V_s) to a variable $\tilde{V}_s$ choosing the (XOR like) mapping so that V_s can be recovered from $(\tilde{V}_s,U)$ while $\tilde{V}_s$ alone is independent of (X,Y). The hints are $M_1{=}(\tilde{V}_s,V_1)$ and $M_2{=}(U,V_2)$. Since

the tuple $(\tilde{V}_s, V_1)$ takes value in the set $\mathcal{V}_s \times \mathcal{V}_1$ of size $c_s c_1$ we must have that $c_s c_1 \leq |\mathcal{M}_1|$. Likewise, we must have that $c_s c_2 \leq |\mathcal{M}_2|$. Because c_s, c_1, and c_2 are positive integers, they thus satisfy (9) ahead. Alice does not use randomization if $c_s = 1$.

Theorem 1 (Finite Blocklength Guessing Version): For every triple $(c_s, c_1, c_2) \in \mathbb{N}^3$ satisfying

$$c_s \leq |\mathcal{M}_1| \wedge |\mathcal{M}_2|\,,\; c_1 \leq \lfloor |\mathcal{M}_1|/c_s \rfloor\,,\; c_2 \leq \lfloor |\mathcal{M}_2|/c_s \rfloor\,, \quad (9)$$

there is a choice of the conditional PMF in (1) for which Bob's ambiguity about X is upper-bounded by

$$\mathscr{A}_{\mathrm{B}}^{(\mathrm{g})}(P_{X,Y}) < 1 + 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log(c_s c_1 c_2) + 1)}, \quad (10)$$

and Eve's ambiguity about X is lower-bounded by

$$\mathscr{A}_{\mathrm{E}}(P_{X,Y}) \geq (1 + \ln|\mathcal{X}|)^{-\rho}\, 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log(c_1 + c_2))}. \quad (11)$$

Conversely, for every conditional PMF, Bob's ambiguity is lower-bounded by

$$\mathscr{A}_{\mathrm{B}}^{(\mathrm{g})}(P_{X,Y}) \geq (1+\ln|\mathcal{X}|)^{-\rho} 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log|\mathcal{M}_1||\mathcal{M}_2|)} \vee 1, \quad (12)$$

and Eve's ambiguity is upper-bounded by

$$\mathscr{A}_{\mathrm{E}}(P_{X,Y}) \leq (|\mathcal{M}_1|^{\rho} \wedge |\mathcal{M}_2|^{\rho}) \mathscr{A}_{\mathrm{B}}^{(\mathrm{g})}(P_{X,Y}) \wedge 2^{\rho H_{\tilde{\rho}}(X|Y)}. \quad (13)$$

Theorem 2 (Finite Blocklength List Version): If $|\mathcal{M}_1||\mathcal{M}_2| > \log|\mathcal{X}| + 2$, then for every triple $(c_s, c_1, c_2) \in \mathbb{N}^3$ satisfying

$$c_s \leq |\mathcal{M}_1| \wedge |\mathcal{M}_2|\,,\; c_1 \leq \lfloor |\mathcal{M}_1|/c_s \rfloor\,,\; c_2 \leq \lfloor |\mathcal{M}_2|/c_s \rfloor\,, \quad (14a)$$

$$c_s c_1 c_2 > \log|\mathcal{X}| + 2, \quad (14b)$$

there is a choice of the conditional PMF in (1) for which Bob's ambiguity about X is upper-bounded by

$$\mathscr{A}_{\mathrm{B}}^{(\mathrm{l})}(P_{X,Y}) < 1 + 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log(c_s c_1 c_2 - \log|\mathcal{X}| - 2) + 2)}, \quad (15)$$

and Eve's ambiguity about X is lower-bounded by

$$\mathscr{A}_{\mathrm{E}}(P_{X,Y}) \geq (1 + \ln|\mathcal{X}|)^{-\rho}\, 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log(c_1 + c_2))}. \quad (16)$$

Conversely, for every conditional PMF, Bob's ambiguity is lower-bounded by

$$\mathscr{A}_{\mathrm{B}}^{(\mathrm{l})}(P_{X,Y}) \geq 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log|\mathcal{M}_1||\mathcal{M}_2|)} \vee 1, \quad (17)$$

and Eve's ambiguity is upper-bounded by

$$\mathscr{A}_{\mathrm{E}}(P_{X,Y}) \leq (|\mathcal{M}_1|^{\rho} \wedge |\mathcal{M}_2|^{\rho}) \mathscr{A}_{\mathrm{B}}^{(\mathrm{l})}(P_{X,Y}) \wedge 2^{\rho H_{\tilde{\rho}}(X|Y)}. \quad (18)$$

We sketch a proof of Theorems 1 and 2 in Section V ahead. Here, we discuss an important implication of the theorems:

Note 3: Let $\mathscr{U}_{\mathrm{B}} \leq 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log|\mathcal{M}_1||\mathcal{M}_2|)} \vee 1$. There is a choice of the conditional PMF in (1) so that, neglecting polylogarithmic factors of $|\mathcal{X}|$, Bob's ambiguity satisfies the upper bound $\mathscr{U}_{\mathrm{B}}$, while Eve's ambiguity is guaranteed to be $(|\mathcal{M}_1|^{\rho} \wedge |\mathcal{M}_2|^{\rho}) \mathscr{U}_B \wedge 2^{\rho H_{\tilde{\rho}}(X|Y)}$.

To show that Note 3 holds, we next argue that the bounds in Theorems 1 and 2 are tight in the sense that with a judicious choice of (c_s, c_1, c_2) the achievability results (namely (10)–(11) in the "guessing version" and (15)–(16) in the "list version") match the corresponding converse results (namely (12)–(13) in the "guessing version" and (17)–(18) in the "list version") up to polylogarithmic factors of $|\mathcal{X}|$. By possibly relabeling the hints, we can assume w.l.g. that $|\mathcal{M}_2| \leq |\mathcal{M}_1|$. If $|\mathcal{M}_2|$ exceeds $2^{H_{\tilde{\rho}}(X|Y)}$ we can choose $(c_s, c_1, c_2) = (|\mathcal{M}_2|, 1, 1)$. Neglecting polylogarithmic factors of $|\mathcal{X}|$, this choice guarantees that Bob's ambiguity be close to one, while Eve's ambiguity is $2^{\rho H_{\tilde{\rho}}(X|Y)}$. Suppose now that $|\mathcal{M}_2|$ does not exceed $2^{H_{\tilde{\rho}}(X|Y)}$. In this case we can choose (c_s, c_1, c_2) so that $c_1 \geq c_2$ and, neglecting logarithmic factors of $|\mathcal{X}|$, so that $c_s c_2 = |\mathcal{M}_2|$ while $c_s c_1 c_2$ assumes any given integer value between $|\mathcal{M}_2|$ and $(|\mathcal{M}_1|\,|\mathcal{M}_2|) \wedge 2^{H_{\tilde{\rho}}(X|Y)}$. This indeed proves the claim: neglecting polylogarithmic factors of $|\mathcal{X}|$, we can guarantee that Bob's ambiguity satisfy any given upper bound no smaller than the RHS of (12) or (17), while Eve's ambiguity satisfies (13) or (18) with equality.

B. Asymptotic Results

Consider now the asymptotic regime where (X, Y) is an n-tuple. In this case the results are the same for both versions of the problem, and we thus refer to both $\mathscr{A}_{\mathrm{B}}^{(\mathrm{g})}$ and $\mathscr{A}_{\mathrm{B}}^{(\mathrm{l})}$ by $\mathscr{A}_{\mathrm{B}}$. With a judicious choice of (c_s, c_1, c_2) one can show that Theorems 1 and 2 imply the following asymptotic result:

Corollary 4: Let $\{(X_i, Y_i)\}_{i \in \mathbb{N}}$ be a discrete-time stochastic process with finite alphabet $\mathcal{X} \times \mathcal{Y}$, and suppose its conditional Rényi entropy-rate $H_{\tilde{\rho}}(\boldsymbol{X}|\boldsymbol{Y})$ is well-defined. Given any positive rate-pair (R_1, R_2), the privacy-exponent is

$$\overline{E_{\mathrm{E}}} = \begin{cases} \rho(R_1 \wedge R_2 \wedge H_{\tilde{\rho}}(\boldsymbol{X}|\boldsymbol{Y})), & R_1 + R_2 > H_{\tilde{\rho}}(\boldsymbol{X}|\boldsymbol{Y}) \\ -\infty, & R_1 + R_2 < H_{\tilde{\rho}}(\boldsymbol{X}|\boldsymbol{Y})\,. \end{cases} \quad (19)$$

In the full version of this paper [1], we generalize Corollary 4 to a scenario where Bob's ambiguity may grow exponentially with a given normalized (by n) exponent E_{B}.

IV. Lists and Guesses

The results for the "guessing version" and the "list version" are remarkably similar. To understand why, we relate task-encoders to guessing functions. We show that a good guessing function induces a good task-encoder and vice versa:

Theorem 5: Let (X, Y) be drawn from the finite set $\mathcal{X} \times \mathcal{Y}$ according to the PMF $P_{X,Y}$. Using the side-information Y, a stochastic task-encoder describes the task X by a chance variable M, which it draws from a finite set $\mathcal{M}$ according to some conditional PMF

$$\mathbb{P}[M = m | X = x, Y = y]\,,\; m \in \mathcal{M},\; x \in \mathcal{X},\; y \in \mathcal{Y}. \quad (20)$$

For any PMF (20) define for all $m \in \mathcal{M}$ and $y \in \mathcal{Y}$ the lists

$$\mathcal{L}_m^y = \{x \in \mathcal{X} \colon \mathbb{P}[X = x | Y = y, M = m] > 0\}\,. \quad (21)$$

1) For every conditional PMF (20) the lists $\{\mathcal{L}_m^y\}$ induce a guessing function $G(\cdot|Y)$ for X such that

$$\mathbb{E}[G(X|Y)^{\rho}] \leq |\mathcal{M}|^{\rho}\, \mathbb{E}\big[|\mathcal{L}_M^Y|^{\rho}\big]. \quad (22)$$

2) Every guessing function $G(\cdot|Y)$ for X and every positive integer $v \leq |\mathcal{X}|$ satisfying

$$|\mathcal{M}| \geq v\,(\lfloor \log\lceil |\mathcal{X}|/v \rceil \rfloor + 1) \quad (23)$$

induce a 0-1 valued conditional PMF (20)—i.e., a deterministic task-encoder—whose lists $\{\mathcal{L}_m^y\}$ satisfy

$$\mathbb{E}\big[|\mathcal{L}_M^Y|^\rho\big] \le \mathbb{E}[\lceil G(X|Y)\,/v\rceil^\rho]\,. \tag{24}$$

To prove Theorem 5, we need the following fact:

Fact 6: Fix a positive integer u, and let $h(\cdot)$ map every $k \in \{1,\dots,u\}$ to $\lfloor \log k \rfloor$. Then,

$$\big|\big\{\tilde{k} \in \{1,\dots,u\} : h(\tilde{k}) = h(k)\big\}\big| \le k,\; k \in \{1,\dots,u\}\,. \tag{25}$$

Proof: If $k, \tilde{k} \in \{1,\dots,u\}$ are such that $h(\tilde{k}) = h(k)$, then $2^{\lfloor \log k \rfloor} \le \tilde{k} < 2^{\lfloor \log k \rfloor + 1}$. Hence, (25) holds. ∎

Proof of Theorem 5: As to the first part, suppose we are given a conditional PMF (20) with corresponding lists $\{\mathcal{L}_m^y\}$ as in (21). For each $y \in \mathcal{Y}$, order the lists $\{\mathcal{L}_m^y\}_{m\in\mathcal{M}}$ in increasing order of their cardinalities, and order the elements in each list in some arbitrary way. Now consider the guessing order where we first guess the elements of the first (and smallest) list in their respective order followed by those elements in the second list that have not yet been guessed (i.e., that are not contained in the first list) and we continue until concluding by guessing those elements of the last (and longest) list that have not been previously guessed. Let $G(\cdot|Y)$ be the corresponding guessing function, and observe that

$$\begin{aligned}\mathbb{E}[G(X|Y)^\rho] &= \sum_{x,y} P_{X,Y}(x,y)\,|\{\tilde{x}\colon G(\tilde{x}|y) \le G(x|y)\}|^\rho \\ &\overset{(a)}{\le} \sum_{x,y} P_{X,Y}(x,y)\,|\mathcal{M}|^\rho \min_{m\colon x\in\mathcal{L}_m^y} |\mathcal{L}_m^y|^\rho \\ &\le |\mathcal{M}|^\rho\,\mathbb{E}\big[|\mathcal{L}_M^Y|^\rho\big],\end{aligned}$$

where (a) holds because for all $x, \tilde{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$ a necessary condition for $G(\tilde{x}|y) \le G(x|y)$ is that $\tilde{x} \in \mathcal{L}_{\tilde{m}}^y$ for some $\tilde{m} \in \mathcal{M}$ satisfying $|\mathcal{L}_{\tilde{m}}^y| \le \min_{m\colon x\in\mathcal{L}_m^y} |\mathcal{L}_m^y|$, and the number of lists whose size does not exceed $\min_{m\colon x\in\mathcal{L}_m^y} |\mathcal{L}_m^y|$ is at most $|\mathcal{M}|$.

As to the second part, suppose we are given a guessing function $G(\cdot|Y)$ for X and a positive integer $v \le |\mathcal{X}|$ that satisfies (23). Let $\mathcal{Z} = \{0,\dots,v-1\}$ and $\mathcal{S} = \{0,\dots,\lfloor \log\lceil |\mathcal{X}|\,/v\rceil\rfloor\}$. From (23) it follows that $|\mathcal{M}| \ge |\mathcal{Z}|\,|\mathcal{S}|$. It thus suffices to prove the existence of a task-encoder that uses only $|\mathcal{Z}|\,|\mathcal{S}|$ possible descriptions, and we thus assume w.l.g. that $\mathcal{M} = \mathcal{Z}\times\mathcal{S}$. That is, using the side-information y the task-encoder (deterministically) describes x by $m = (z,s)$. The encoding involves two steps:

Step 1: In Step 1 the encoder first computes $Z \in \mathcal{Z}$ as the remainder of the Euclidean division of $G(X|Y) - 1$ by $|\mathcal{Z}|$. It then constructs from $G(\cdot|Y)$ a guessing function $G(\cdot|Y,Z)$ for X as follows. Given $Y = y$ and $Z = z$, the task X must be in the set $\mathcal{X}_{y,z} \triangleq \{x \in \mathcal{X}\colon (G(x|y) - 1) \equiv z \mod |\mathcal{Z}|\}$. The encoder constructs the guessing function $G(\cdot|y,z)$ so that—in the corresponding guessing order—we first guess the elements of $\mathcal{X}_{y,z}$ in increasing order of $G(x|y)$. For $l \in \{1,\dots,|\mathcal{X}_{y,z}|\}$ our l-th guess x_l is thus the element of $\mathcal{X}_{y,z}$ for which $G(x_l|y) = z + 1 + (l-1)\,|\mathcal{Z}|$. Once we have guessed all the elements of $\mathcal{X}_{y,z}$ we guess the remaining elements of $\mathcal{X}$ in some arbitrary order. This order is immaterial because X is guaranteed to be in the set $\mathcal{X}_{y,z}$. Since $z+1 \in \{1,\dots,|\mathcal{Z}|\}$ we find that $G(x|y,z) = \lceil G(x|y)\,/\,|\mathcal{Z}|\rceil$ whenever $x \in \mathcal{X}_{y,z}$. But X is guaranteed to be in the set $\mathcal{X}_{y,z}$. Hence, the guessing function $G(\cdot|Y,Z)$ for X satisfies

$$G(X|Y,Z) = \lceil G(X|Y)\,/\,|\mathcal{Z}|\rceil\,. \tag{26}$$

Step 2: In Step 2 the encoder first computes $S = \lfloor \log G(X|Y,Z)\rfloor \in \mathcal{S}$, and then describes the task X by $M \triangleq (Z,S)$. Given $Y = y$, $Z = z$, and $S = s$, the task X must be in the set $\mathcal{X}_s^{y,z} \triangleq \{x \in \mathcal{X}\colon \lfloor \log G(x|y,z)\rfloor = s\}$. Fact 6 and the fact that the guessing function $G(\cdot|y,z)$ is a bijection imply that $|\mathcal{X}_s^{y,z}| \le G(x|y,z)$ for all $x \in \mathcal{X}_s^{y,z}$. Since X is guaranteed to be in the set $\mathcal{X}_s^{y,z}$ we have $|\mathcal{X}_S^{Y,Z}| \le G(X|Y,Z)$. From $M = (Z,S)$ and (21) we obtain that the list $\mathcal{L}_M^Y$ is contained in the set $\mathcal{X}_S^{Y,Z}$ and thus satisfies

$$|\mathcal{L}_M^Y| \le G(X|Y,Z)\,. \tag{27}$$

Recalling that $|\mathcal{Z}| = v$ we conclude from (26) and (27)

$$\mathbb{E}\big[|\mathcal{L}_M^Y|^\rho\big] \le \mathbb{E}[G(X|Y,Z)^\rho] = \mathbb{E}[\lceil G(X|Y)\,/v\rceil^\rho]\,. \tag{28}$$

Since Z and S are deterministic given (X,Y) the conditional PMF (20) associated with $M = (Z,S)$ is 0-1 valued. ∎

The choice of v as $\lfloor |\mathcal{M}|\,/\,(\lfloor \log|\mathcal{X}|\rfloor + 1)\rfloor$ and [5, Equation (26)], i.e., $\lceil \xi\rceil^\rho < 1 + 2^\rho\xi^\rho, \xi \ge 0$, imply our next result:

Corollary 7: Every guessing function $G(\cdot|Y)$ for X induces a deterministic task-encoder corresponding to a 0-1 valued conditional PMF (20) that satisfies

$$\mathbb{E}\big[|\mathcal{L}_M^Y|^\rho\big] \le 1 + 2^\rho\,\mathbb{E}[G(X|Y)^\rho]\left(\frac{|\mathcal{M}|}{\log|\mathcal{X}|+1} - 1\right)^{-\rho}\,. \tag{29}$$

Combined with Arikan's bounds [4, Theorem 1 and Proposition 4] on $\mathbb{E}[G(X|Y)^\rho]$, Equations (22) and (29) provide an upper and a lower bound on the smallest $\mathbb{E}\big[|\mathcal{L}_M^Y|^\rho\big]$ that is achievable for a given $|\mathcal{M}|$. These bounds are weaker than [5, Theorem 1.1 and Theorem 6.1] in the finite blocklength regime but tight enough to prove the asymptotic results [5, Theorem 1.2 and Theorem 6.2].

Another interesting corollary to Theorem 5 results from the choice of v as 1:

Corollary 8: For $|\mathcal{M}| = \lfloor \log|\mathcal{X}|\rfloor + 1$ every guessing function $G(\cdot|Y)$ induces a deterministic task-encoder for which

$$\mathbb{E}\big[|\mathcal{L}_M^Y|^\rho\big] \le \mathbb{E}\big[G(X|Y)^\rho\big]. \tag{30}$$

The corollary can be used to show that the results for the "guessing version" and the "list version" differ only by polylogarithmic factors of $|\mathcal{X}|$ [1].

V. On the Proof of Theorems 1 and 2

To prove Theorems 1 and 2, we must quantify how additional side-information Z helps guessing. We show that if Z takes value in a finite set $\mathcal{Z}$, then it can reduce the ρ-th moment of the number of guesses by at most a factor of $|\mathcal{Z}|^{-\rho}$.

Lemma 9: Let (X,Y,Z) be drawn from the finite set $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ according to the PMF $P_{X,Y,Z}$. Then,

$$\mathbb{E}[G^*(X|Y,Z)^\rho] \ge \mathbb{E}\big[\lceil G^*(X|Y)\,/\,|\mathcal{Z}|\rceil^\rho\big], \tag{31}$$

where $G^*(\cdot|Y,Z)$ minimizes $\mathbb{E}\big[G(X|Y,Z)^\rho\big]$ and $G^*(\cdot|Y)$ minimizes $\mathbb{E}\big[G(X|Y)^\rho\big]$. Equality holds whenever $Z = f(X,Y)$ for some $f\colon \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ for which $f(x,y)=f(\tilde{x},y)$ implies either $\lceil G^*(x|y)\,/\,|\mathcal{Z}|\rceil \neq \lceil G^*(\tilde{x}|y)\,/\,|\mathcal{Z}|\rceil$ or $x=\tilde{x}$. Such a function always exists because for all $l\in\mathbb{N}$ at most $|\mathcal{Z}|$ different $x\in\mathcal{X}$ satisfy $\lceil G^*(x|y)\,/\,|\mathcal{Z}|\rceil = l$.

Proof: If $g(x,y)\in \arg\min_{z\in\mathcal{Z}} G^*(x|y,z)$ for $(x,y)\in\mathcal{X}\times\mathcal{Y}$, then $\mathbb{E}[G^*(X|Y,Z)^\rho]\geq \min_G \mathbb{E}[G(X|Y,g(X,Y))^\rho]$. It thus suffices to prove (31) for the case where Z is deterministic given (X,Y), and we thus assume w.l.g. that $Z=g(X,Y)$ for some function $g\colon \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$. Consider

$$\mathbb{E}[G(X|Y,Z)^\rho] = \sum_{x,y} P_{X,Y}(x,y)\, G(x|y,g(x,y))^\rho\,, \quad (32)$$

where $G(\cdot|Y,g(X,Y))$ is a guessing function. Note that $G(x|y,g(x,y))=G(\tilde{x}|y,g(\tilde{x},y))$ implies $g(x,y)\neq g(\tilde{x},y)$ for all $y\in\mathcal{Y}$ and distinct $x,\,\tilde{x}\in\mathcal{X}$. For every $l\in\mathbb{N}$ there are thus at most $|\mathcal{Z}|$ different $x\in\mathcal{X}$ for which $G(x|y,g(x,y))=l$. For each $y\in\mathcal{Y}$ order the possible realizations of X in decreasing order of $P_{X,Y}(x,y)$, i.e., in decreasing order of their posterior probabilities given $Y=y$, and let x_j^y denote the j-th element. Clearly, (32) is minimum over $g(\cdot,\cdot)$ and $G(\cdot|Y,g(X,Y))$ if for $l\in\mathbb{N}$ and $y\in\mathcal{Y}$ we have $G(x|y,g(x,y))=l$ whenever $x=x_j^y$ for some j satisfying $(l-1)\,|\mathcal{Z}|+1\leq j\leq l\,|\mathcal{Z}|$ or, equivalently, $\lceil j/\,|\mathcal{Z}|\rceil = l$. Since $G^*(\cdot|Y)$ minimizes $\mathbb{E}\big[G(X|Y)^\rho\big]$, it orders the elements of $\mathcal{X}$ in decreasing order of their posterior probabilities given Y. We can thus choose x_j^y to be the unique $x\in\mathcal{X}$ for which $G^*(x|y)=j$. Hence, (32) is minimized if $f(\cdot,\cdot)$ satisfies the specifications in the lemma, $g(\cdot,\cdot)=f(\cdot,\cdot)$, and $G(x|y,f(x,y))=\lceil G^*(x|y)\,/\,|\mathcal{Z}|\rceil$. The minimum equals the RHS of (31). ∎

Lemma 9 and [5, Equation (26)] imply the following result:

Corollary 10: Let (X,Y) be drawn from the finite set $\mathcal{X}\times\mathcal{Y}$ according to the PMF $P_{X,Y}$, and let $\mathcal{Z}$ be a finite set. There exists a function $f\colon \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ such that for $Z=f(X,Y)$

$$\min_G \mathbb{E}[G(X|Y,Z)^\rho] < 1+2^\rho\,|\mathcal{Z}|^{-\rho}\min_G \mathbb{E}[G(X|Y)^\rho]\,. \quad (33)$$

Conversely, every chance variable Z with alphabet $\mathcal{Z}$ satisfies

$$\min_G \mathbb{E}[G(X|Y,Z)^\rho] \geq |\mathcal{Z}|^{-\rho}\min_G \mathbb{E}\big[G(X|Y)^\rho\big] \vee 1. \quad (34)$$

We now sketch the proofs of Theorems 1 and 2 starting with the direct part. Fix $(c_s,c_1,c_2)\in\mathbb{N}^3$ satisfying (9) in the "guessing version" and (14) in the "list version". For each $\nu\in\{s,1,2\}$ let V_ν be a chance variable taking value in the set $\mathcal{V}_\nu=\{0,\dots,c_\nu-1\}$. Corollary 10 and [4, Proposition 4] imply that there is a 0-1 valued conditional PMF $\mathbb{P}[(V_s,V_1,V_2)=(v_s,v_1,v_2)\,|X=x,Y=y]$ for which

$$\min_G \mathbb{E}\big[G(X|Y,V_s,V_1,V_2)^\rho\big] < 1+2^{\rho(H_{\tilde{\rho}}(X|Y)-\log(c_sc_1c_2)+1)}, \quad (35)$$

and likewise [5, Theorem 6.1] implies that there is a 0-1 valued conditional PMF for which

$$\mathbb{E}\big[|\mathcal{L}^Y_{V_s,V_1,V_2}|^\rho\big] < 1+2^{\rho(H_{\tilde{\rho}}(X|Y)-\log(c_sc_1c_2-\log|\mathcal{X}|-2)+2)}. \quad (36)$$

Both (9) and (14) imply $|\mathcal{M}_1|\geq c_sc_1$ and $|\mathcal{M}_2|\geq c_sc_2$. It thus suffices to prove (10)–(11) and (15)–(16) for a conditional PMF (1) that assigns positive mass only to c_sc_1 elements of $\mathcal{M}_1$ and c_sc_2 elements of $\mathcal{M}_2$, and we thus assume w.l.g. that $\mathcal{M}_1=\mathcal{V}_s\times\mathcal{V}_1$ and $\mathcal{M}_2=\mathcal{V}_s\times\mathcal{V}_2$. Hence, we can choose $M_1=(V_s\oplus_{c_s}U,V_1)$ and $M_2=(U,V_2)$, where U is independent of (X,Y) and uniformly distributed over $\mathcal{V}_s$, and where $\oplus_{c_s}$ denotes modulo-c_s addition. For this choice (10) follows from (35) and (15) from (36). The proof of (11) and (16) is more involved. It builds on the following two ideas: 1) Since U is computable from both (X,M_1) and (X,M_2) we can w.l.g. assume that Eve must guess (X,U) instead of X. 2) Given two guessing functions $G_1(\cdot,\cdot\,|Y,M_1)$ and $G_2(\cdot,\cdot\,|Y,M_2)$ for (X,U), one can show that $G_1(\cdot,\cdot\,|Y,M_1)\wedge G_2(\cdot,\cdot\,|Y,M_2)$ behaves like a guessing function $G(\cdot,\cdot|Y,Z)$ for (X,U), where the additional side-information Z assumes at most $c_s(c_1+c_2)$ different values. Once 1) and 2) have been established, the proof is concluded by Corollary 10, [4, Theorem 1], and $H_{\tilde{\rho}}(X,U|Y)=H_{\frac{1}{1+\rho}}(X\,|Y)+\log c_s$.

The converse is straightforward: In the "guessing version" the bound (12) on Bob's ambiguity follows from Corollary 10 and [4, Theorem 1]. In the "list version" (17) follows from [5, Theorem 6.1] and the observation that $\mathscr{A}_{\mathrm{B}}^{(l)}$ is minimized if the PMF in (1) is 0-1 valued. Clearly, Eve's ambiguity satisfies $\mathscr{A}_{\mathrm{E}}(P_{X,Y}) \leq \min_{k\in\{1,2\}}\Big(\min_{G_k}\mathbb{E}[G_k(X|Y,M_k)^\rho]\Big)$, and Corollary 10 implies for each $k\in\{1,2\}$ and $l\in\{1,2\}\setminus\{k\}$

$$\min_G \mathbb{E}[G(X|Y,M_1,M_2)^\rho] \geq |\mathcal{M}_l|^{-\rho}\min_{G_k}\mathbb{E}[G_k(X|Y,M_k)^\rho]\,.$$

Since $\min_G \mathbb{E}[G(X|Y,M_1,M_2)^\rho]\leq \mathbb{E}\big[|\mathcal{L}^Y_{M_1,M_2}|^\rho\big]$ we thus find that in both versions Eve's ambiguity exceeds Bob's by at most a factor of $|\mathcal{M}_1|^\rho\wedge|\mathcal{M}_2|^\rho$. Due to [4, Proposition 4] and because Eve can guess X using only Y we have $\min_G\mathbb{E}[G(X|Y)^\rho]\leq 2^{\rho H_{\tilde{\rho}}(X|Y)}$. Hence, (13) and (18) hold.

VI. Extensions

In the full version of this paper [1], we discuss several modifications of the model: We extend the analysis to a scenario where Eve observes only the first (less secure) hint and to a setting where Alice produces one (public) hint and encrypts it using a secret key, which is available to Bob but not to Eve. We also generalize the asymptotic results to the case where Bob and Eve must reconstruct X^n within a given distortion D (cf. [5], [7]).

References

[1] A. Bracher, E. Hof, and A. Lapidoth, "Secrecy constrained encoding for guessing decoders and list-decoders," in preparation.

[2] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper," *IEEE Trans. Inf. Theory*, IT-45, No. 6, pp. 1860–1866, Sep. 1999.

[3] J. L. Massey, "Guessing and entropy," *Proc. of IEEE Int. Symp. on Inf. Theory (ISIT)*, p. 204, Jun. 1994.

[4] E. Arikan, "An inequality on guessing and its applications to sequential decoding," *IEEE Trans. Inf. Theory*, IT-42, No. 1, pp. 99–105, Jan. 1996.

[5] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *arXiv:1401.6338 [cs.IT]*, 2014.

[6] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," *Topics in Inf. Theory*, Vol. 17, No. 6, pp. 41-52, 1977.

[7] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, IT-44, No. 3, pp. 1041–1056, May 1998.