Evaluating Multiple Guesses by an Adversary via a Tunable Loss Function

Gowtham R. Kurri, Oliver Kosut, Lalitha Sankar

***Abstract*—We consider a problem of guessing, wherein an adversary is interested in knowing the value of the realization of a discrete random variable X on observing another correlated random variable Y. The adversary can make multiple (say, k) guesses. The adversary's guessing strategy is assumed to minimize α-loss, a class of tunable loss functions parameterized by α. It has been shown before that this loss function captures well known loss functions including the exponential loss ($\alpha = 1/2$), the log-loss ($\alpha = 1$) and the 0-1 loss ($\alpha = \infty$). We completely characterize the optimal adversarial strategy and the resulting expected α-loss, thereby recovering known results for $\alpha = \infty$. We define an information leakage measure from the k-guesses setup and derive a condition under which the leakage is unchanged from a single guess.**

I. INTRODUCTION

The classical guessing problem involves an adversary interested in finding the value of a realization of a discrete random variable X by asking a series of questions in an adaptive manner until an affirmative answer is received. A commonly used performance metric for the guessing problem is the expected number of guesses required until X is guessed correctly, or more generally a moment of this number. Massey [1] established a lower bound on the expected number of guesses in terms of the entropy of X. Later, Arikan [2] investigated the problem of bounding the moments of the number of guesses in terms of the Rényi entropy [3] of X. Further connections between Rényi entropy and guessing are explored in [4]–[7].

We study the guessing problem where an adversary makes a fixed number of guesses. Such a setting finds applications in several practical scenarios. For example, an adversary is allowed several guesses to login with a password before getting locked-out. We consider a setup where an adversary is interested in guessing the unknown value of a random variable X on observing another correlated random variable Y, where X and Y are jointly distributed according to P_{XY} over the finite support $\mathcal{X} \times \mathcal{Y}$. Since the adversary makes a fixed number of guesses k, we focus on evaluating the adversary's success using loss functions that in turn can measure the information leaked by Y about X. To this end, we model the adversary's strategy using α-loss, a class of tunable loss functions parameterized by $\alpha \in (0, \infty]$ [8], [9]. This class captures the well-known exponential loss ($\alpha = 1/2$) [10], log-loss ($\alpha = 1$) [11]–[13], and the 0-1 loss ($\alpha = \infty$) [12], [14]. The adversary then seeks to find the optimal (possibly randomized) guessing strategy that minimizes the expected α-loss over k guesses.

Devising *guessing strategies* with the quest to optimize certain performance metrics of an adversary has several applications in information theory and related fields; this includes sequential decoding [2], guessing codewords [15], botnet attacks [5], [7], to name a few. In [5], the authors consider a guessing problem with a fixed number of guesses allowing for randomized guessing strategies (similar to our setting) and analyze the exponential behaviour of the probability of success in guessing the sequences. A closely related work is that of *maximal leakage* [16] which captures the information leaked when an adversary maximizes its probability of correctly guessing (equivalent to minimizing 0-1 loss) an unknown function of X; they further generalize this notion to k-guesses, and they show the resulting leakage measure is unchanged.

Our main contributions are as follows:

- We completely characterize the minimal expected α-loss for k guesses (Theorem 1), thereby recovering known results for $\alpha = \infty$ [16]. To the best of our knowledge, such a result even for log-loss ($\alpha = 1$) under multiple guesses was not explored earlier. We derive a technique for transforming the optimization problem over the probability simplex associated with multiple random variables to that of with a single random variable using tools drawn from duality in linear programming, which may be of independent interest (Lemma 2).
- We define a measure of information leakage for k guesses of an adversary motivated by α-leakage [8, Definition 5] and show that it does not change with the number of guesses for a class of probability distributions P_{XY} (Theorem 2).

II. BACKGROUND AND PROBLEM DEFINITION

We first review α-loss and then define the minimal expected α-loss for k guesses. Later, we define a measure of information leakage based on this.

Definition 1 (α-loss [8], [9])**.** *For $\alpha \in (0,1) \cup (1,\infty)$, the α-loss is a function defined from $[0,1]$ to $\mathbb{R}_+$ as*

$$\ell_\alpha(p) := \frac{\alpha}{\alpha - 1}\left(1 - p^{\frac{\alpha-1}{\alpha}}\right). \tag{1}$$

The authors are with the School of Electrical, Computer and Energy Engineering at Arizona State University. Email: gkurri@asu.edu, okosut@asu.edu, lsankar@asu.edu

This work is supported in part by NSF grants CIF-1901243, CIF-1815361, and CIF-2007688.

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517733

It is defined by continuous extension for $\alpha = 1$ and $\alpha = \infty$, respectively, and is given by

$$\ell_1(p) = \log \frac{1}{p},\ \ell_\infty(p) = 1 - p. \tag{2}$$

Notice that $\ell_\alpha(p)$ is decreasing in p.

Definition 2 (Minimal expected α-loss for k guesses)**.** *Consider random variables $(X, Y) \sim P_{XY}$ and an adversary that makes k guesses $\hat{X}_{[1:k]} = \hat{X}_1, \hat{X}_2, \ldots, \hat{X}_k$ on observing Y such that $X - Y - \hat{X}_{[1:k]}$ is a Markov chain. Let $P_{\hat{X}_{[1:k]}|Y}$ be a strategy for estimating X from Y in k guesses. For $\alpha \in (0, \infty]$, the minimal expected α-loss for k guesses is defined as*

$$\begin{aligned}&\mathcal{ME}_\alpha^{(k)}(P_{XY})\\ &:= \min_{P_{\hat{X}_{[1:k]}|Y}} \sum_{x,y} P_{XY}(x,y)\ell_\alpha\left(\mathrm{P}\left(\bigcup_{i=1}^{k}(\hat{X}_i = x|Y = y)\right)\right).\end{aligned} \tag{3}$$

We interpret $\mathrm{P}\left(\bigcup_{i=1}^{k}(\hat{X}_i = x)|Y = y\right)$ as the probability of correctly estimating $X = x$ given $Y = y$ in k guesses. An adversary seeks to find the optimal guessing strategy in (3). Note that the optimization problem in (3) was solved for a special case of $k = 1$ by Liao *et al.* [8, Lemma 1]. Notice that

$$\mathcal{ME}_\alpha^{(k)}(P_{XY}) = \sum_y P_Y(y)\mathcal{ME}_\alpha^{(k)}(P_{X|Y=y}), \tag{4}$$

where we have slightly abused the notation in the R.H.S. of (4). Hence, in view of (4), in order to solve the optimization problem in (3), it suffices to solve for a case where $Y = \emptyset$, i.e.,

$$\mathcal{ME}_\alpha^{(k)}(P_X) := \min_{P_{\hat{X}_{[1:k]}}} \sum_x P_X(x)\ell_\alpha\left(P\left(\bigcup_{i=1}^{k}(\hat{X}_i = x)\right)\right). \tag{5}$$

Also, in the sequel, it suffices to consider the optimization problem in (5) only for the case where $k < n$, where P_X is supported on $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ because if $k \geq n$, we have $\mathcal{ME}_\alpha^{(k)}(P_X) = 0$, since a strategy $P^*_{\hat{X}_{[1:k]}}$ such that $P^*_{\hat{X}_{[1:n]}}(x_1, x_2, \ldots, x_n) = 1$ is optimal.

Motivated by α-leakage [8, Definition 5] which captures how much information an adversary can learn about a random variable X from a correlated random variable Y when a single guess is allowed, we define a leakage measure which captures the information an adversary can learn when k guesses are allowed. This definition is also related to maximal leakage under k guesses [16].

Definition 3 (α-leakage with k guesses)**.** *Given a joint distribution P_{XY} and k estimators $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_k$ with the same support as X, the α-leakage from X to Y with k guesses is defined as*

$$\begin{aligned}&\mathcal{L}_\alpha^{(k)}(X \to Y) \triangleq\\ &\frac{\alpha}{\alpha - 1}\log\frac{\max\limits_{P_{\hat{X}_{[1:k]}|Y}} \mathbb{E}\left[\mathrm{P}\left(\bigcup\limits_{i=1}^{k}(\hat{X}_i = X)|Y\right)^{\frac{\alpha-1}{\alpha}}\right]}{\max\limits_{P_{\hat{X}_{[1:k]}}} \mathbb{E}\left[\mathrm{P}\left(\bigcup\limits_{i=1}^{k}(\hat{X}_i = X)\right)^{\frac{\alpha-1}{\alpha}}\right]},\end{aligned} \tag{6}$$

for $\alpha \in (0, 1) \cup (1, \infty)$.

III. Main Results

Theorem 1 (Minimal expected α-loss for k guesses)**.** *Consider a P_X supported on $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ such that $p_1 \geq p_2 \geq \cdots \geq p_n$, where $p_i := P_X(x_i)$, for $i \in [1 : n]$. Then the minimal expected α-loss for k guesses is given by*

$$\mathcal{ME}_\alpha^{(k)}(P_X) = \frac{\alpha}{\alpha - 1}\sum_{i=s^*}^{n} p_i\left(1 - \left(\frac{(k - s^* + 1)p_i^\alpha}{\sum_{j=s^*}^{n} p_j^\alpha}\right)^{\frac{\alpha-1}{\alpha}}\right), \tag{7}$$

where

$$s^* = \min\left\{r \in \{1, 2, \ldots, k\} : \frac{(k - r + 1)p_r^\alpha}{\sum_{i=r}^{n} p_i^\alpha} \leq 1\right\}. \tag{8}$$

Remark 1. It can be inferred from Theorem 1 that in the optimal guessing strategy, the adversary guesses the s^* most likely outcomes, and uses an updated tilted distribution on the rest of the outcomes (see also (28)). For the special case when $k = 2$ and $s^* = 1$, this optimal strategy is exactly the same as that of a seemingly different guessing problem considered in [17, Section II-B].

Remark 2. Notice that whenever $s^* = 1$ in (8), the expression in (7) simplifies to

$$\frac{\alpha}{\alpha - 1}\left(1 - k^{\frac{\alpha-1}{\alpha}}\exp\left(\frac{1 - \alpha}{\alpha}H_\alpha(X)\right)\right), \tag{9}$$

where $H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n} p_i^\alpha\right)$ is the Rényi entropy of order α [3]. Also, note that for the special case of $k = 1$, we always have $s^* = 1$, thereby recovering [8, Lemma 1].

Corollary 1 (Minimal expected log-loss $\{\alpha = 1\}$ for k guesses)**.** *Under the notations of Theorem 1, the minimal expected log-loss for k guesses is given by*

$$\begin{aligned}\mathcal{ME}_1^{(k)}(P_X) = H(X) - H_{s^*}&\left(p_1, p_2, \ldots, p_{s^*-1}, \sum_{i=s^*}^{n} p_i\right)\\ &- \left(\sum_{i=s^*}^{n} p_i\right)\log(k - s^* + 1),\end{aligned} \tag{10}$$

where $s^* = \min\left\{r \in \{1, 2, \ldots, k\} : \frac{(k-r+1)p_r}{\sum_{i=r}^{n} p_i} \leq 1\right\}$ and $H_{s^*}(q_1, q_2, \ldots, q_{s^*}) := \sum_{i=1}^{s^*} q_i \log\frac{1}{q_i}$ is the entropy function.

Corollary 2 (Minimal expected 0-1 loss $\{\alpha = \infty\}$ for k guesses)**.** *Under the notations of Theorem 1, the minimal expected* 0-1 *loss for* k *guesses is given by*

$$\begin{aligned}\mathcal{ME}_{\infty}^{(k)}(P_X) &= 1 - \sum_{i=1}^{k} p_i \\ &= 1 - \max_{\substack{a_1,a_2,\ldots,a_k: \\ a_l \neq a_m, l \neq m}} \sum_{i=1}^{k} P_X(a_i). \end{aligned} \tag{11}$$

The following theorem shows the robustness of α-leakage to the number of guesses for a class of probability distributions P_{XY}. Let $P_{X|Y=y}^{(\alpha)}$ denote the tilted distribution of $P_{X|Y=y}$, i.e., $P_{X|Y}^{(\alpha)}(x|y) = \frac{P_{X|Y}(x|y)^{\alpha}}{\sum_x P_{X|Y}(x|y)^{\alpha}}$.

Theorem 2 (Robustness of α-leakage to number of guesses)**.** *Consider a* P_{XY} *such that* $P_{X|Y}^{(\alpha)}(x|y) \leq \frac{1}{k}$*, for all* x, y *and* $P_X^{(\alpha)}(x) \leq \frac{1}{k}$*, for all* x*. Then*

$$\mathcal{L}_{\alpha}^{(k)} = \mathcal{L}_{\alpha}^{(1)}. \tag{12}$$

The proofs of Theorems 1 and 2 are given in the following section.

IV. Proofs of Main Results

We begin with the following lemmas which will be useful in the proof of Theorem 1. It is intuitive to expect that an optimal strategy, $P^*_{\hat{X}_{[1:k]}}$, puts zero weight on ordered tuples $(a_1, a_2, \ldots, a_k)$ (denoted as $a_{[1:k]}$ in the sequel) whenever $a_i = a_j$ for some $i \neq j$, since there is no advantage in guessing the same estimate more than once. The following lemma based on the monotonicity of the α-loss formalizes this.

Lemma 1. *If* $P^*_{\hat{X}_{[1:k]}}$ *is an optimal strategy for the optimization problem in* (5)*, then*

$$P^*_{\hat{X}_{[1:k]}}(a_{[1:k]}) = 0, \text{ for all } a_{[1:k]} \text{ s.t. } a_i = a_j, \text{ for some } i \neq j.$$

The proof of Lemma 1 is deferred to Appendix A.

Remark 3. An important consequence of Lemma 1 is that, if $P^*_{\hat{X}_{[1:k]}}$ is an optimal strategy for the optimization problem in (5), then we have

$$\sum_x \mathrm{P}^* \left(\bigcup_{i=1}^{k} (\hat{X}_i = x) \right) = k, \tag{13}$$

where the probability P^* is taken with respect to an optimal strategy $P^*_{\hat{X}_{[1:k]}}$. Hence, it suffices to consider the optimization in (5) over all the strategies $P_{\hat{X}_{[1:k]}}$ satisfying (13).

Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ be the support of P_X. A vector $(t_1, t_2, \ldots, t_n)$ such that $\sum_{i=1}^n t_i = k$ is said to be *admissible* if there exists a strategy $P_{\hat{X}_{[1:k]}}$ satisfying

$$t_i = \mathrm{P} \left(\bigcup_{j=1}^{k} (\hat{X}_j = x_i) \right), \text{ for all } i \in [1:n]. \tag{14}$$

Equivalently, (14) can be written as the following system of linear equations.

$$t_i = \sum_{a_{[1:k]}: \bigcup_{j=1}^{k} (a_j = x_i)} P_{\hat{X}_{[1:k]}}(a_{[1:k]}), \text{ for all } i \in [1:n]. \tag{15}$$

In general, in order to determine whether a vector $(t_1, t_2, \ldots, t_n)$ is admissible or not, we need to solve a linear programming problem (LPP) with number of variables and constraints that are polynomial in the support size of P_X, i.e, n. Nonetheless, the following lemma based on Farkas' lemma [18, Proposition 6.4.3] completely characterizes the necessary and sufficient conditions for the admissibility of a vector $(t_1, t_2, \ldots, t_n)$.

Lemma 2. *A vector* $(t_1, t_2, \ldots, t_n)$ *such that* $\sum_{i=1}^n t_i = k$ *is admissible if and only if* $0 \leq t_i \leq 1$*, for all* $i \in [1:n]$.

The proof of Lemma 2 is deferred to Appendix B. We are now ready to prove Theorem 1.

Proof of Theorem 1. From the definition of the minimal expected α-loss for k guesses in (5), we have

$$\begin{aligned} &\mathcal{ME}_{\alpha}^{(k)}(P_X) \\ &= \min_{P_{\hat{X}_{[1:k]}}} \frac{\alpha}{\alpha - 1} \left[\sum_{i=1}^{n} p_i \left(1 - \mathrm{P} \left(\bigcup_{j=1}^{k} (\hat{X}_j = x_i) \right)^{\frac{\alpha-1}{\alpha}} \right) \right] \qquad (16) \\ &= \min_{P_{\hat{X}_{[1:k]}}} \frac{\alpha}{\alpha - 1} \left[\sum_{i=1}^{n} p_i \left(1 - \mathrm{P} \left(\bigcup_{j=1}^{k} (\hat{X}_j = x_i) \right)^{\frac{\alpha-1}{\alpha}} \right) \right] \\ &\qquad \text{s.t. } \sum_{i=1}^{n} \mathrm{P} \left(\bigcup_{j=1}^{k} (\hat{X}_j = x_i) \right) = k \qquad (17) \\ &= \min_{t_1,\ldots,t_n} \frac{\alpha}{\alpha - 1} \left[\sum_{i=1}^{n} p_i (1 - t_i^{\frac{\alpha-1}{\alpha}}) \right] \\ &\qquad \text{s.t. } \sum_{i=1}^{n} t_i = k, \\ &\qquad 0 \leq t_i \leq 1, \ i \in [1:n], \qquad (18) \end{aligned}$$

where (17) follows from Lemma 1 and Remark 3, and (18) follows from the change of variable $t_i = \mathrm{P} \left(\bigcup_{j=1}^{k} (\hat{X}_j = x_i) \right)$ and Lemma 2. Consider the Lagrangian

$$\begin{aligned} \mathcal{L} = \frac{\alpha}{\alpha - 1} \left[\sum_{i=1}^{n} p_i (1 - t_i^{\frac{\alpha-1}{\alpha}}) \right] + \lambda \left(\sum_{i=1}^{n} t_i - k \right) \\ + \sum_{i=1}^{n} \mu_i (t_i - 1) \end{aligned} \tag{19}$$

The Karush-Kuhn-Tucker (KKT) conditions [19, Chapter 5.5.3] are given by

$$\text{(Stationarity): } \frac{\partial \mathcal{L}}{\partial t_i} = 0, i \in [1:n],$$

$$\text{i.e., } t_i = \left(\frac{p_i}{\lambda + \mu_i}\right)^{\alpha}, i \in [1:n], \tag{20}$$

$$\text{(Primal feasibility): } \sum_{i=1}^{n} t_i = k, 0 \le t_i \le 1, i \in [1:n], \tag{21}$$

$$\text{(Dual feasibility): } \mu_i \ge 0, i \in [1:n], \tag{22}$$

$$\text{(Complementary slackness): } \mu_i(t_i - 1) = 0, i \in [1:n]. \tag{23}$$

Notice that for $\alpha > 1$, $t^{\frac{\alpha-1}{\alpha}}$ is a concave function of t, meaning the overall objective function in (18) is convex. For $\alpha < 1$, $t^{\frac{\alpha-1}{\alpha}}$ is a convex function of t, but since $\frac{\alpha}{\alpha-1}$ is negative, the overall function is again convex. Thus (18) amounts to a convex optimization problem. Now since KKT conditions are necessary and sufficient conditions for optimality in a convex optimization problem, it suffices to find values of t_i, $i \in [1:n]$, λ, μ_i, $i \in [1:n]$ satisfying (20)–(23) in order to solve the optimization problem (18).

First we simplify the KKT conditions (20)–(23) in the following manner.

- For i such that $\left(\frac{p_i}{\lambda}\right)^{\alpha} \le 1$, we take $\mu_i = 0$ and $t_i = \left(\frac{p_i}{\lambda}\right)^{\alpha}$.
- For i such that $\left(\frac{p_i}{\lambda}\right)^{\alpha} > 1$, we take $\mu_i = p_i - \lambda$ and $t_i = 1$. Notice that for such i, we have $\mu_i > 0$, since $p_i > \lambda$.

This is equivalent to choosing $t_i = \min\left\{\left(\frac{p_i}{\lambda}\right)^{\alpha}, 1\right\}$ and $\mu_i = 0$ or $\mu_i = p_i - \lambda$ depending on whether $t_i = \left(\frac{p_i}{\lambda}\right)^{\alpha}$ or $t_i = 1$, respectively, for each $i \in [1:n]$. Notice that this choice is consistent with the KKT conditions (20)–(23) except for that λ has to be chosen appropriately satisfying $\sum_{i=1}^{n} t_i = k$ also. In effect, we have essentially reduced the KKT conditions (20)–(23) to the following equations by eliminating μ_i's:

$$t_i = \min\left\{\left(\frac{p_i}{\lambda}\right)^{\alpha}, 1\right\}, i \in [1:n], \tag{24}$$

$$\sum_{i=1}^{n} t_i = k. \tag{25}$$

We solve the equations (24) and (25) by considering the following k mutually exclusive and exhaustive cases (clarified later) based on P_X.

Case 1 $\left(\frac{p_1^{\alpha}}{\sum_{i=1}^{n} p_i^{\alpha}} \le \frac{1}{k}\right)$:
Consider the choice

$$\lambda = \left(\frac{\sum_{i=1}^{n} p_i^{\alpha}}{k}\right)^{\frac{1}{\alpha}}, \; t_i = \frac{k p_i^{\alpha}}{\sum_{j=1}^{n} p_j^{\alpha}}, i \in [1:n]. \tag{26}$$

This choice satisfies (24) and (25) since $\frac{k p_1^{\alpha}}{\sum_{i=1}^{n} p_i^{\alpha}} \le 1$ and $p_1 \ge p_2 \cdots \ge p_n$.

Case 's' ($2 \le s \le k$) $\left(\frac{(k-s+2)p_{s-1}^{\alpha}}{\sum_{i=s-1}^{n} p_i^{\alpha}} > 1, \frac{(k-s+1)p_s^{\alpha}}{\sum_{i=s}^{n} p_i^{\alpha}} \le 1\right)$:
Consider the choice

$$\lambda = \left(\frac{\sum_{i=s}^{n} p_i^{\alpha}}{k-s+1}\right)^{\frac{1}{\alpha}}, \tag{27}$$

$$t_i = 1, i \in [1:s-1], t_i = \frac{(k-s+1)p_i^{\alpha}}{\sum_{j=s}^{n} p_j^{\alpha}}, i \in [s:n]. \tag{28}$$

This choice satisfies (24)

- for $i \in [1:s-1]$ because $\frac{(k-s+2)p_{s-1}^{\alpha}}{\sum_{i=s-1}^{n} p_i^{\alpha}} > 1$ and $p_1 \ge p_2 \ge \cdots \ge p_{s-1}$, and
- for $i \in [s:n]$ because $\frac{(k-s+1)p_s^{\alpha}}{\sum_{i=s}^{n} p_i^{\alpha}} \le 1$ and $p_s \ge p_{s+1} \ge \cdots \ge p_n$.

Also, this choice clearly satisfies (25). Finally, notice that the condition for Case 's', $2 \le s \le n$, can be written as

$$\frac{(k-i+1)p_i^{\alpha}}{\sum_{j=i}^{n} p_j^{\alpha}} > 1, \text{ for } i \in [1:s-1], \frac{(k-s+1)p_s^{\alpha}}{\sum_{i=s}^{n} p_i^{\alpha}} \le 1 \tag{29}$$

since $\frac{(k-s+2)p_{s-1}^{\alpha}}{\sum_{i=s-1}^{n} p_i^{\alpha}} > 1$ and $p_1 \ge p_2 \ge \cdots \ge p_{s-1}$. This proves that the cases considered above are mutually exclusive and exhaustive, and together with the case-wise analysis gives the expression for the minimal expected α-loss for k guesses as presented in Theorem 1. □

The proof of Corollary 1 follows by taking limit $\alpha \to 1$ using L'Hôpital's rule in the result of Theorem 1 and rearranging the terms. The proof of Corollary 2 follows by taking limit $\alpha \to \infty$ in Theorem 1.

Proof of Theorem 2. From the definition of α-leakage with k guesses in (6), we have

$$\begin{aligned}
&\mathcal{L}_{\alpha}^{(k)}(X \to Y) \\
&= \frac{\alpha}{\alpha-1} \log \frac{\max\limits_{P_{\hat{X}_{[1:k]}|Y}} \mathbb{E}\left[\mathrm{P}\left(\bigcup_{i=1}^{k} (\hat{X}_i = X)|Y\right)^{\frac{\alpha-1}{\alpha}}\right]}{\max\limits_{P_{\hat{X}_{[1:k]}}} \mathbb{E}\left[\mathrm{P}\left(\bigcup_{i=1}^{k} (\hat{X}_i = X)\right)^{\frac{\alpha-1}{\alpha}}\right]} && (30)\\
&= \frac{\alpha}{\alpha-1} \log \frac{k^{\frac{\alpha-1}{\alpha}} \exp\left(\frac{1-\alpha}{\alpha} H_{\alpha}^{A}(X|Y)\right)}{k^{\frac{\alpha-1}{\alpha}} \exp\left(\frac{1-\alpha}{\alpha} H_{\alpha}(X)\right)} && (31)\\
&= \frac{\alpha}{\alpha-1} \log \frac{\exp\left(\frac{1-\alpha}{\alpha} H_{\alpha}^{A}(X|Y)\right)}{\exp\left(\frac{1-\alpha}{\alpha} H_{\alpha}(X)\right)} && (32)\\
&= \mathcal{L}_{\alpha}^{(1)}, && (33)
\end{aligned}$$

where (31) follows from Theorem 1, in particular from the case when $s^* = 1$ since $P_{X|Y}^{(\alpha)}(x|y) \le \frac{1}{k}$, for all x, y and $P_X^{(\alpha)}(x) \le \frac{1}{k}$, for all x, and $H_{\alpha}^{A}(X|Y)$ in (31) is the Arimoto conditional entropy [20] defined as $H_{\alpha}^{A}(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_{y} \left(\sum_{x} P_{XY}(x,y)^{\alpha}\right)^{\frac{1}{\alpha}}$. □

V. Conclusion

There are many questions to be further studied. For example, analogously to maximal leakage [16] and maximal α-leakage [8], we can define a maximal version of α-leakage with k guesses. As shown in [16], for $\alpha = \infty$, this quantity does not change with k; it would be interesting to understand whether this is also true for other α.

Appendix A
Proof of Lemma 1

Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and $P_X(x_i) = p_i$, for $i \in [1:n]$. Consider $a_{[1:k]}$ such that $a_i = a_j$ for some $i \neq j$. There exists a $b_{[1:k]}$ such that for each $i \in [1:k]$, we have $a_i = b_j$ for some j and $b_r \neq a_j$ for some r and any j. Consider

$$\frac{\alpha}{\alpha-1}\left[\sum_{i=1}^{n} p_i \left(1 - \mathrm{P}^* \left(\bigcup_{j=1}^{k} (\hat{X}_j = x_i)\right)^{\frac{\alpha-1}{\alpha}}\right)\right]. \tag{34}$$

Let $\mathcal{A}$ and $\mathcal{B}$ denote the sets of all multiset permutations of $a_{[1:k]}$ and $b_{[1:k]}$, respectively, when $a_{[1:k]}$ and $b_{[1:k]}$ are treated as multisets. Let $q_{a_1,a_2,\ldots,a_k} := \sum_{r_{[1:k]} \in \mathcal{A}} P_{\hat{X}_{[1:k]}}(r_{[1:k]})$ and $q_{b_1,b_2,\ldots,b_k} := \sum_{r_{[1:k]} \in \mathcal{B}} P_{\hat{X}_{[1:k]}}(r_{[1:k]})$. Each term out of the n terms in (34) will either contain both $q_{a_{[1:k]}}$ and $q_{b_{[1:k]}}$ (say, type 1), contain just $q_{b_{[1:k]}}$ alone (say, type 2), or does not contain both (say, type 3). We now construct a new strategy $P_{\hat{X}_{[1:k]}}$ by incorporating the value of $q_{a_{[1:k]}}$ into $q_{b_{[1:k]}}$ making the value of new $q_{a_{[1:k]}}$ equal to zero. Now the values of the terms of type 2 strictly decrease as the α-loss function is strictly decreasing in its argument while retaining the values of the terms of types 1 and 3. This leads to a contradiction since $P^*_{X_{[1:k]}}$ is assumed to be an optimal strategy. So, $P_{\hat{X}_{[1:k]}}(a_{[1:k]}) = 0$. Repeating the same argument as above for all such $a_{[1:k]}$ s.t. $a_i = a_j$, for some $i \neq j$ completes the proof.

Appendix B
Proof of Lemma 2

'Only if' part: Suppose a vector $(t_1, t_2, \ldots, t_n)$ is admissible. Then there exists $P_{\hat{X}_{[1:k]}}$ satisfying (15). Using (14), since t_i is probability of a certain event, we have

$$0 \leq t_i \leq 1, \text{ for } i \in [1:n].$$

'If' part: Suppose $0 \leq t_i \leq 1$, for $i \in [1:n]$. Summing up all the equations in (15) over $i \in [1:n]$ and using $\sum_{i=1}^{n} t_i = k$, we get

$$P_{\hat{X}_{[1:k]}}(a_{[1:k]}) = 0, \text{ for all } a_{[1:k]} \text{ s.t. } a_i = a_j, \text{ for some } i \neq j.$$

With this, (15) can be written in the form of system of linear equation only in terms of non-negative variables of the form

$$q_{i_1,i_2,\ldots,i_k} := \sum_{\sigma \in S_n} P_{\hat{X}_{[1:k]}}(x_{i_{\sigma(1)}}, x_{i_{\sigma(2)}}, \ldots, x_{i_{\sigma(n)}}), \tag{35}$$

where $i_1, i_2, \ldots, i_k$ are all distinct and belong to $[1:n]$. Here the sum is computed over all the permutations σ of the set $\{1, 2, \ldots, n\}$. The set of all such permutations is denoted by S_n. With this, the system of equations in (15) can be written in the form $AQ = b$, $Q \geq 0$. Here A is a $n \times \binom{n}{k}$-matrix, where the rows are indexed by $i \in [1:n]$ and columns are indexed by $(i_1, i_2, \ldots, i_k)$, where $i_1, i_2, \ldots, i_k$ are all distinct and belong to $[1:n]$. In particular, in the column indexed by $(i_1, i_2, \ldots, i_k)$, the entry of A corresponding to i_j^{th} row is 1, for $j \in [1:k]$. All the remaining entries of the matrix A are zeros. Q is $\binom{n}{k}$-length vector of variables of the form $q_{i_1,i_2,\ldots,i_k}$. b is an n-length vector with $b_i = t_i$. We are interested in the feasibility of the system $AQ = b$, $Q \geq 0$. We use the Farkas' lemma [18, Proposition 6.4.3] in linear programming for checking this. It states that the system $AQ = b$ has a non-negative solution if and only if every $y \in \mathbb{R}^n$ with $y^\top A \geq 0$ also implies $y^\top b \geq 0$. For our problem, $y^\top A \geq 0$ is equivalent to

$$\sum_{j=1}^{k} y_{i_j} \geq 0, \text{ for all distinct } i_1, i_2, \ldots, i_k \in [1:n]. \tag{36}$$

Without loss of generality, let us assume that $y_i \leq y_{i+1}$, $i \in [1:n-1]$. Then (36) is equivalent to

$$\sum_{i=1}^{k} y_i \geq 0. \tag{37}$$

Now consider

$$\begin{aligned}
&\sum_{i=1}^{n} y_i t_i \\
&= \sum_{i=1}^{k} y_i t_i + y_{k+1} t_{k+1} + \sum_{i=k+2}^{n} y_i t_i && (38)\\
&= \sum_{i=1}^{k} y_i + \sum_{i=1}^{k} y_i (t_i - 1) + y_{k+1} t_{k+1} + \sum_{i=k+2}^{n} y_i t_i && (39)\\
&\geq \sum_{i=1}^{k} y_i + y_{k+1} \sum_{i=1}^{k} (t_i - 1) + y_{k+1} t_{k+1} + \sum_{i=k+2}^{n} y_i t_i && (40)\\
&\geq \sum_{i=1}^{k} y_i + y_{k+1} \sum_{i=1}^{k} (t_i - 1) + y_{k+1} t_{k+1} + y_{k+1} \sum_{i=k+2}^{n} t_i && (41)\\
&= \sum_{i=1}^{k} y_i + y_{k+1} \left(\sum_{i=1}^{n} t_i - k\right) && (42)\\
&= \sum_{i=1}^{k} y_i && (43)\\
&\geq 0, && (44)
\end{aligned}$$

where (40) follows because $y_i \leq y_{k+1}$ and $t_i - 1 \leq 0$, for $i \in [1:k]$, (41) follows because $y_i \geq y_{k+1}$, for $i \in [k+2:n]$, and (43) follows because $\sum_{i=1}^{n} t_i = k$, (44) follows from (37). Now using the Farkas' lemma, $AQ = b$, has a non-negative solution, i.e., the vector $(t_1, t_2, \ldots, t_n)$ is admissible.

REFERENCES

[1] J. L. Massey, "Guessing and entropy," in *IEEE International Symposium on Information Theory*, 1994, p. 204.

[2] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[3] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 1961, pp. 547–561.

[4] I. Sason and S. Verdú, "Improved bounds on lossless source coding and guessing moments via Rényi measures," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4323–4346, 2018.

[5] S. Salamatian, W. Huleihel, A. Beirami, A. Cohen, and M. Médard, "Why botnets work: Distributed brute-force attacks need no synchronization," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 9, pp. 2288–2299, 2019.

[6] M. M. Christiansen and K. R. Duffy, "Guesswork, large deviations, and Shannon entropy," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 796–802, 2013.

[7] N. Merhav and A. Cohen, "Universal randomized guessing with application to asynchronous decentralized brute–force attacks," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 114–129, 2020.

[8] J. Liao, O. Kosut, L. Sankar, and F. P. Calmon, "Tunable measures for information leakage and applications to privacy-utility tradeoffs," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 8043–8066, 2019.

[9] T. Sypherd, M. Diaz, L. Sankar, and P. Kairouz, "A tunable loss function for binary classification," in *IEEE International Symposium on Information Theory*, 2019, pp. 2479–2483.

[10] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[11] N. Merhav and M. Feder, "Universal prediction," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2124–2147, 1998.

[12] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On surrogate loss functions and f-divergences," *The Annals of Statistics*, vol. 37, no. 2, pp. 876–904, 2009.

[13] T. A. Courtade and R. D. Wesel, "Multiterminal source coding with an entropy-based distortion measure," in *IEEE International Symposium on Information Theory*, 2011, pp. 2040–2044.

[14] P. L. Bartlett, M. I. Jordan, and J. D. Mcauliffe, "Convexity, classification, and risk bounds," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 138–156, 2006.

[15] C. E. Pfister and W. G. Sullivan, "Renyi entropy, guesswork moments, and large deviations," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2794–2800, 2004.

[16] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1625–1657, 2020.

[17] W. Huleihel, S. Salamatian, and M. Médard, "Guessing with limited memory," in *2017 IEEE International Symposium on Information Theory*, 2017, pp. 2253–2257.

[18] J. Matousek and B. Gartner, *Understanding and Using Linear Programming*. Springer, 2007.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[20] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," *Topics in information theory*, 1977.