On Renyi Entropies and their Applications to Guessing Attacks in Cryptography

Serdar Boztaş
RMIT University
Melbourne, Australia
E-mail:serdar.boztas@rmit.edu.au
Preprint of paper with same title
IEICE Transactions on Fundamentals
97-A(12):2014

September 2014

Abstract

We consider single and multiple attacker scenarios in guessing and obtain bounds on various success parameters in terms of Renyi entropies. We also obtain a new derivation of the union bound.

Keywords: Guessing, Renyi Entropy, Shannon Entropy, Predictability, Brute force attacks, Cryptography, Work factor.

1 Introduction

Let X be an unknown discrete random variable $X \in \mathcal{X}$ with $\mathcal{X}$ finite or countable, and with distribution $\mathbb{P}$. This random variable could, for example, represent an unknown key for a cryptosystem, or an unknown password. In practice, the *guessor* is not all-powerful and can only ask atomic questions (e.g., query keys/passwords) regarding singletons in $\mathcal{X}$. We assume that a sequence of questions of the form **Is** $X = x$**?** are posed until the first YES answer determines the value of the random variable X.

The problem of guessing has been investigated in the context of sequential decoding [1] and source-channel coding ([2, 3, 12]) as well as in security applications ([15, 6, 7, 11]). We provide an overview of previous work in the last subsection. It is also possible to define guessing in the presence of distortion or source uncertainty ([2, 12, 16]) but we do not pursue this any further here.

While it is attractive to have a number of different guessors working in parallel in trying to obtain the value of the random variable X, there are also some pitfalls in making this approach flexible, in terms of participants entering and leaving the group performing the attack and in terms of partitioning the search space $\mathcal{X}$. This can turn out to be complicated, since it is quite likely that the computational power of each participant (thus the rate at which they can implement the guessing mechanism) can vary a great deal. These factors make the study of oblivious distributed guessing of interest.

However the model we are considering in this paper is also relevant to distributed attacks mounted from independent nodes on the internet. As a generic example, consider that independent attackers are probing different nodes with password guesses, with the condition that the overall password distribution of the probed notes all come from some probability distribution $\mathbb{P}$ over the password search space. The goal of the attackers may be maximizing the probability of success that *at least one of them* finds the password, or minimizing the probability that *the password is not found* after some fixed number of guesses, etc. Some of the results in this paper have been partially presented in [5], but significant new results have also been obtained, namely Theorems 2, 3 and 4.

1.1 Definitions and Notation

We provide some definitions in this section. A *guessing strategy* for identifying X is a procedure for generating successive questions of the above type until a YES answer is obtained. Formally, any such procedure can be represented by a function $G : \mathcal{X} \to \{1, 2, \ldots\}$ where $G(k)$ equals the time index of the question **Is** $X = k$**?**.

Note that functions G corresponding to valid guessing strategies cannot be totally arbitrary. Clearly, G must be invertible on its range $\{1, 2, \ldots\}$ since only one element may be probed at any given time. Moreover, since we are assuming the answers to the queries **Is** $X = k$**?** are noiseless, it is enough to consider guessing strategies which ask the above question *exactly*

once for each $k \geq 1$. This formally corresponds to the mapping G being one-to-one and onto. Any function satisfying these two conditions will be called a *guessing function*. Every guessing function defines a valid guessing strategy and conversely.

Assuming that the guessor knows the probability distribution $\mathbb{P}$ (otherwise see [16]), she is interested in minimizing the number of questions required to determine X. This goal can be formalised in a number of ways, such as minimizing a positive moment $\mathbb{E}[G^\rho]$ (mostly $\rho = 1$ is of interest) where

$$\mathbb{E}[G^\rho] = \sum_{x \in \mathcal{X}} \mathbb{P}(x) G(x)^\rho = \sum_{k \geq 1} k^\rho \mathbb{P}(G^{-1}(k)).$$

The Rényi entropy of order α of X is a generalization of the Shannon entropy defined by

$$H_\alpha(X) = \frac{\log\left(\sum_{X \in \mathcal{Y}} \mathbb{P}(X)^\alpha\right)}{1-\alpha} \qquad \alpha \in [0,1) \cup (1,\infty),$$

and obeys $\lim_{\alpha \to 1} H_\alpha(X) = H(X)$ as well as being strictly decreasing in α unless the Y is uniform on its support.

Sometimes, a related random variable $Y \in \mathcal{Y}$ with some joint distribution $\mathbb{P}(X,Y)$ is available to the guessor, who then proceeds to guess the possible values of X as above. In this case we use the notation

$$\mathbb{E}[G(X|Y)] = \sum_{y \in \mathcal{Y}} \mathbb{P}(Y) \mathbb{E}[G(X|Y=y)]$$

for the expectation of a conditional guessing strategy.

1.2 Previous Work and Novelty of Our Results

Here we give a brief overview of previous work that is most relevant to the focus of this paper.

The link between guessing and entropy was popularized by Massey [8]. The problem of bounding the expected number of guesses in terms of Renyi entropies was investigated by Arikan in the context of sequential decoding [1]. Pliam independently investigated the relationship between entropy, guesswork and security [15]. Feder and Merhav considered the relationship between predictability and entropy [10]. Pfister and Sullivan [14] and Malone and Sullivan [9] also obtained results related to guesswork and entropies.

The problem of a cipher with a guessing wiretapper, and the problem of guessing subject to distortion was investigated by Merhav and Arikan [2]. The problem of guessing under source uncertainty was investigated by Sundaresan [16]. Hanawal and Sundaresan have also unified the work on guessing exponents, using large deviations theory, in [18]. The problem of multilevel guessing was investigated by Merhav, Roth and Arikan [12].

This paper obtains some new bounds on guessing attacks under novel scenarios not previously considered, as well as discussing and comparing with existing results by the author and others.

The paper is organised as follows. In Section 2, we state the problem we are considering and define and analyze the guessing algorithm which minimizes the expected number of guesses, in a distributed environment for the case of a single attacker. Section 3 discusses the case of multiple attackers. Section 4 concludes the paper.

2 Single Attacker Case

2.1 Single Unconstrained Attacker

Brute force predictability uses the idea of guessing every value of X one by one in order of decreasing probability, when the distribution $\mathbb{P}(x)$ is known. Previous results in this domain have been obtained by the author in [4] and by Pliam in [15]. In [4] the following bound was proved.

Theorem 1. *The expected number of guesses,* for a user with the optimal guessing sequence*,* $\mathbb{E}[G]$ *obeys the upper bound*

$$\mathbb{E}[G] \leq \frac{1}{2}\left[\sum_{k=1}^{N}\sqrt{p_k}\right]^2 + \frac{1}{2} = \frac{2^{H_{1/2}(X)+1}}{2}, \tag{1}$$

2.1.1 Renyi Entropy of Some Unique Order Exactly Determines Expected Number of Guesses

In this section we assume for notational simplicity that the optimal guessing sequence is $G(x_k) = k$, for $k = 1, \ldots, M$. This is equivalent to the condition

$$\mathbb{P}(x_k) \geq \mathbb{P}(x_{k+1}), \qquad k \geq 1. \tag{2}$$

With X as in the previous subsection, we first prove the bound below and note that this bound can directly be obtained from the distribution of X:

Theorem 2. *For any random variable X with finite range $\mathcal{X}$ the expected number of guesses in the optimal guessing sequence satisfies*

$$\ln \mathbb{E}[G(X)] \leq (1-\alpha^*) H_{\alpha^*}(X)$$

where $\alpha^* = \min\{\alpha_k \mid k = 1, 2, \ldots M\}$, and where

$$\alpha_k = \frac{\ln k}{\ln \mathbb{P}(x_k)} + 1,$$

and $M = |\mathcal{X}|$. Further, $\alpha^* = 0$ if and only if the distribution $\mathbb{P}$ is uniform on $\mathcal{X}$, otherwise we have $\alpha^* \in (0, 1)$. This bound can also be written as

$$\mathbb{E}[G(X)] \leq \sum_{k=1}^{M} \mathbb{P}(x_k)^{\alpha^*}.$$

Proof: Consider the expectation $\mathbb{E}[G(X)]$ and note that

$$\begin{aligned} \mathbb{E}[G(X)] &= \sum_{k=1}^{M} k \cdot \mathbb{P}(x_k) \\ &\leq \sum_{k=1}^{M} \mathbb{P}(x_k)^{\alpha_k} \\ \text{provided} \qquad & k \cdot \mathbb{P}(x_k) \leq \mathbb{P}(x_k)^{\alpha_k} \end{aligned}$$

which can be ensured with equality if we choose α_k such that

$$k = \mathbb{P}(x_k)^{\alpha_k - 1} \qquad \text{or} \qquad \alpha_k = \frac{\ln k}{\ln \mathbb{P}(x_k)} + 1,$$

for $k = 1, 2, \ldots, M$.

The extremal distribution within the class of distributions on M points satisfying (1) which yields the maximum expected value $\mathbb{E}[G(X)]$ is the uniform distribution with $\mathbb{P}(x_k) = 1/M$ for $k = 1, 2, \ldots, M$. For this distribution, $\alpha_M = 0$ and the upper bound yields $\mathbb{E}[G(X)] \leq M$ while we know that $\mathbb{E}[G(X)] = (1 + M)/2$.

We claim that all the other distributions on M points satisfying (1) yield

$$\alpha_1 = 1, \qquad \alpha_k \in (0,1) \ \text{ for } \ k \geq 2.$$

To prove the claim, consider an arbitrary distribution on M points satisfying (1) and assume that $\alpha^* = 0$. This would imply that there exists a k such that

$$\alpha_k = 0 \ \Leftrightarrow \ \frac{\ln k}{\ln \mathbb{P}(x_k)} = -1 \ \Leftrightarrow \ \ln k = -\ln \mathbb{P}(x_k)$$

$$\Leftrightarrow \ \mathbb{P}(x_k) = \frac{1}{k}.$$

Since we're considering distributions satisfying (1), the above equations imply that we have the uniform distribution. We have hence proved that $\alpha^* = 0$ **if and only if** $\mathbb{P}$ is the uniform distribution. Moreover, it is obvious that $\alpha_k < 1$ holds since $\mathbb{P}(x_k) \in (0,1)$ ($\Rightarrow \ln \mathbb{P}(x_k) < 0$) for nontrivial distributions on $M > 1$ points. Now assume that we have $\alpha^* < 0$, for some distribution on M points. This would imply that there is a k such that

$$\frac{\ln k}{\ln \mathbb{P}(x_k)} < -1 \ \Leftrightarrow \ \ln k > -\ln \mathbb{P}(x_k)$$

$$\Leftrightarrow \ k > \frac{1}{\mathbb{P}(x_k)} \ \Leftrightarrow \ \frac{1}{k} < \mathbb{P}(x_k).$$

Equation (1) would then imply that $\mathbb{P}(x_m) > \frac{1}{k}$ for $m < k$, or that

$$\sum_{l=1}^{M} \mathbb{P}(x_l) \geq \sum_{l=1}^{k} \mathbb{P}(x_l) > k \cdot (1/k) = 1,$$

which is a contradiction. We have thus proved that $\alpha^* \in [0,1)$ with $\alpha^* = 0$ if and only if $\mathbb{P}$ is uniform. $\square$

Note that this bound is not tight, since it is obtained by matching the two sides of equation

$$\sum_{k=1}^{M} k \cdot \mathbb{P}(x_k) \leq \sum_{k=1}^{M} \mathbb{P}(x_k)^{\alpha^*}$$

term by term. An exact expression for $\mathbb{E}[G(X)]$ in terms of the Rényi entropy can be obtained by solving

$$\sum_{k=1}^{M} k \cdot \mathbb{P}(x_k) = \sum_{k=1}^{M} \mathbb{P}(x_k)^{\alpha^*}$$

for α^* given $\mathbb{P}(\cdot)$. It is not too hard to see that such a unique solution exists since the right hand side is a continuous monotone decreasing function of α^* which ranges between M (when $\alpha^* = 0$) and 1 (when $\alpha = 1$) .The left hand side, i.e., $\mathbb{E}[G(X)]$, is known to lie in $[1, M]$ and hence $\alpha^* \in (0, 1)$. Moreover, we know from Theorem 1 that the bound

$$\sum_{k=1}^{M} k \cdot \mathbb{P}(x_k) \leq \frac{1}{2} \left[\sum_{k=1}^{M} \mathbb{P}(x_k)^{1/2} \right]^2 + \frac{1}{2}$$

holds for a class of distributions including monotone distributions (i.e., optimal guessing). Hence, we can conclude that $\alpha^* \in (0, 1/2)$ will hold (for $M > 2$). We have thus proved

Theorem 3. *The equation*

$$\sum_{k=1}^{M} k \cdot \mathbb{P}(x_k) = \sum_{k=1}^{M} \mathbb{P}(x_k)^{\alpha^*} \tag{3}$$

has a unique solution $\alpha^* \in (0, 1/2)$ and this implies that we have an equality for the expectation $\mathbb{E}[G(X)]$ in terms of the Rényi entropy, viz.

$$\ln \mathbb{E}[G(X)] = (1 - \alpha^*) H_{\alpha^*}(X)$$

Note that while an explicit analytic expression for α^* is likely to remain elusive, given a distribution $\mathbb{P}(\cdot)$, it is computationally simple to compute $\alpha^* \in (0, 1/2)$ to whatever accuracy is desired by, e.g., the bisection method.

2.1.2 A New Derivation of the Union Bound

An important application of the uniform upper bound in section 2 is to use it to derive a version of the union bound for the probability of error. This will be addressed in this section.

We will show that the union bound on the probability of error $\mathbb{P}_{error}$ can be interpreted as a bound on $\mathbb{E}[G(X|Y)] - 1$ where X is the input to a noisy communication channel and Y is its output. Note that it is straightforward

to show that

$$\begin{aligned}
\mathbb{P}_{error} &= \sum_{y=1}^{M} \sum_{m \neq y} \mathbb{P}_{XY}(m, y) \\
1 + \mathbb{P}_{error} &= \sum_{y=1}^{M} \sum_{m \neq y} 2\mathbb{P}_{XY}(m, y) + \sum_{y=1}^{M} \mathbb{P}_{XY}(y, y) \\
&\leq \sum_{y=1}^{M} \sum_{m \neq y} \sigma_y(m) \mathbb{P}_{XY}(m, y) + \sum_{y=1}^{M} \sigma_y(y) \mathbb{P}_{XY}(y, y)
\end{aligned}$$

where $\sigma_y : \{1, \cdots, M\} \to \{1, \cdots, M\}$ is the permutation determined by the conditional guessing scheme $G(\cdot|Y = y)$. This is equivalent to:

$$1 + \mathbb{P}_{error} \leq \mathbb{E}[G(X|Y)].$$

Now we apply the bound in Theorem to the conditional case, by applying the bound to each of the conditional distributions $\mathbb{P}(X|Y = y)$ while y ranges over $\mathcal{Y} = \{1, \ldots, M\}$, to obtain

$$\begin{aligned}
\mathbb{E}[G(X|Y)] &= \sum_{y=1}^{M} \mathbb{P}_Y(y) \mathbb{E}[G(X|Y = y)] \\
&\leq \sum_{y=1}^{M} \mathbb{P}_Y(y) \left\{ \frac{1}{2} \left[\sum_{m=1}^{M} \sqrt{\mathbb{P}_{X|Y}(m|y)} \right]^2 + \frac{1}{2} \right\} \\
&= \frac{1}{2} + \frac{1}{2} \sum_{y=1}^{M} \mathbb{P}_Y(y) \sum_{m,m'=1}^{M} \sqrt{\mathbb{P}_{X|Y}(m|y) \mathbb{P}_{X|Y}(m'|y)}
\end{aligned}$$

which becomes

$$\begin{aligned}
\mathbb{E}[G(X|Y)] &\leq 1 + \frac{1}{2} \sum_{y=1}^{M} \mathbb{P}_Y(y) \sum_{1 \leq m \neq m' \leq M} \sqrt{\mathbb{P}_{X|Y}(m|y) \mathbb{P}_{X|Y}(m'|y)} \\
&= 1 + \frac{1}{2} \sum_{1 \leq m \neq m' \leq M} \sum_{y=1}^{M} \mathbb{P}_Y(y) \sqrt{\mathbb{P}_{X|Y}(m|y) \mathbb{P}_{X|Y}(m'|y)}.
\end{aligned}$$

We therefore obtain (using $\mathbb{P}_{error} \leq \mathbb{E}[G(X|Y)] - 1$) that

$$\begin{aligned}\mathbb{P}_{error} &\leq \frac{1}{2} \sum_{1 \leq m \neq m' \leq M} \left[\sum_{y=1}^{M} \mathbb{P}_Y(y) \sqrt{\mathbb{P}_{X|Y}(m|y)\mathbb{P}_{X|Y}(m'|y)} \right] \\ &= \sum_{m=1}^{M} \sum_{m'=1}^{m-1} \left[\sum_{y=1}^{M} \mathbb{P}_Y(y) \sqrt{\mathbb{P}_{X|Y}(m|y)\mathbb{P}_{X|Y}(m'|y)} \right].\end{aligned}$$

Now, note that if the transmitted symbol is $X = m$, the sum $\sum_{m=1}^{M} \sum_{m'=1}^{m-1}$ can be rewritten $\sum_{m=1}^{M} \sum_{m' \in \mathcal{E}(m)}$ where $\mathcal{E}(m)$ is the "error set" of m. We therefore have

$$\begin{aligned}\mathbb{P}_{error} &\leq \sum_{m=1}^{M} \sum_{m' \in \mathcal{E}(m)} \left[\sum_{y=1}^{M} \mathbb{P}_Y(y) \sqrt{\mathbb{P}_{X|Y}(m|y)\mathbb{P}_{X|Y}(m'|y)} \right] \\ &= \sum_{m=1}^{M} \sum_{m' \in \mathcal{E}(m)} \left[\sum_{y=1}^{M} \sqrt{\mathbb{P}_{XY}(m,y)\mathbb{P}_{XY}(m',y)} \right]\end{aligned}$$

which is a form of the union bound.

Theorem 4. *The union bound of information theory may be written as*

$$\mathbb{P}_{error} \leq \mathbb{E}[G(X|Y)] - 1$$

where Y is the output of a noisy channel which is available to the decoder and X is its input.

2.2 Memory Constrained Guessor Minimizing $\mathbb{E}[G]$

If there are no constraints on time or memory, the optimal strategy is to minimize the expected number of guesses $\mathbb{E}[G]$ and maximize the probability of success at each time index by guessing every point in $\mathcal{X}$ in decreasing order of probability.

We now want to consider the model of an attacker that is constrained, either (a) in memory, or perhaps (b) in power. This leads in case (a) to an attacker who can't keep track of what guesses it has previously made, and in case (b) to an attacker that can only make a predetermined number of guesses. The question we want to answer is the following:

Given $\mathbb{P}(x)$, how should the attacker choose a distribution $\mathbb{Q}(x)$ in order to optimize some performance criterion, when all the attacker does (resp. attackers do) is to draw random sequential guesses from $\mathbb{Q}(x)$?

Let's consider a single attacker who is memory constrained and won't keep track of past guesses, but knows the distribution $\mathbb{P}$ which the opponent uses to draw a single value x from $\mathcal{X}$ according to $\mathbb{P}(x)$.

The guessor generates i.i.d. $X_1, X_2, \ldots,$ from $\mathcal{X}$ according to a distribution $\mathbb{Q}(x)$ again with the goal of minimizing $\mathbb{E}[G]$. Define $G = \min\{k : X_k = X\}$ as a random variable which denotes the number of guesses before she is successful in exposing X. Note that $G = k$ with probability $\sum_{x=1}^{N} \mathbb{P}(x)(1-\mathbb{Q}(x))^{k-1}\mathbb{Q}(x)$, where $k \geq 1$, by a success-fail argument. This is because

$$\begin{aligned}
\mathbb{P}(G=k) &= \sum_{x\in\mathcal{X}} \mathbb{P}(X=x)\mathbb{P}(G=k \mid X=x) \\
&= \sum_{x\in\mathcal{X}} \mathbb{P}(x) Pr\left(\bigcap_{m=1}^{k-1}\{X_m \neq x\} \; and \; X_k = x\right) \\
&= \sum_{x\in\mathcal{X}} \mathbb{P}(x)(1-\mathbb{Q}(x))^{k-1}\mathbb{Q}(x).
\end{aligned}$$

Since the guessing scheme is randomized, it is possible to have an unbounded number of guesses, hence

$$\begin{aligned}
\mathbb{E}[G] &= \sum_{k=1}^{\infty} k \sum_{x\in\mathcal{X}} \mathbb{P}(x)(1-\mathbb{Q}(x))^{k-1}\mathbb{Q}(x) \\
&= \sum_{x\in\mathcal{X}} \mathbb{P}(x)\mathbb{Q}(x) \sum_{k=1}^{\infty} k(1-\mathbb{Q}(x))^{k-1} \\
&= \sum_{x\in\mathcal{X}} \frac{\mathbb{P}(x)\mathbb{Q}(x)}{\mathbb{Q}(x)^2} = \sum_{x\in\mathcal{X}} \frac{\mathbb{P}(x)}{\mathbb{Q}(x)}
\end{aligned}$$

where we used the generating function identity

$$\sum_{k=0}^{\infty} k u^{k-1} = \sum_{k=1}^{\infty} k u^{k-1} = \frac{1}{(1-u)^2}. \tag{4}$$

If we apply Lagrange multipliers with the Lagrangian

$$J = \mathbb{E}[G] + \lambda(\sum_{x\in\mathcal{X}} \mathbb{Q}(x) - 1) = \sum_{x\in\mathcal{X}} \frac{\mathbb{P}(x)}{\mathbb{Q}(x)} + \lambda(\sum_{x\in\mathcal{X}} \mathbb{Q}(x) - 1),$$

we can actually show that $\mathbb{E}[G]$ is minimized when we choose

$$\mathbb{Q}(x) \propto \sqrt{\mathbb{P}(x)},$$

quite an interesting result. This means that the distribution $\mathbb{Q}(x)$ should be "flatter" than $\mathbb{P}(x)$. We have thus proved

Theorem 5. *The distribution which minimizes the expected number of guesses for a random variable with a nontrivial distribution $\mathbb{P}$, i.e., $0 < \mathbb{P}(x) < 1$ for all $x \in \mathcal{X}$, when a single guessor draws her guesses from the distribution $\mathbb{Q}$ is given by*

$$\mathbb{Q}(x) = \frac{\sqrt{\mathbb{P}(x)}}{\sum_{z\in\mathcal{X}} \sqrt{\mathbb{P}(z)}}$$

This can be generalized to the case of multiple independent attackers against one target, or against multiple targets, and we consider the first problem in the rest of this paper.

Note that the Lagrange multipliers ensure that the solution for $\mathbb{Q}(x)$ in Theorem 5 is a minimum, since the expression $\sum_x \mathbb{P}(x)/\mathbb{Q}(x)$ is convex-$\bigcup$ in $(\mathbb{Q}(x_1), \ldots, \mathbb{Q}(x_N))$ if we recall that the $\mathbb{P}(x)$ are given positive constants. Differentiation of J with respect to $\mathbb{Q}(x)$ yields

$$\frac{\partial J}{\partial \mathbb{Q}(x)} = -\frac{\mathbb{P}(x)}{\mathbb{Q}(x)^2} + \lambda = 0, \qquad \forall x \in \mathcal{X}$$

which leads to the theorem above. Furthermore, note that if we choose $\mathbb{Q}(x) = \mathbb{P}(x)$ for all $x \in \mathcal{X}$ which may look like an attractive choice, we obtain $\mathbb{E}[G] = N$ which is unexpectedly high. Let us now consider the optimal value of the expectation which the guessor using Theorem 5 achieves. We obtain

$$\begin{aligned} \mathbb{E}[G] &= \sum_{x\in\mathcal{X}} \frac{\mathbb{P}(x)}{\mathbb{Q}(x)} = \sum_{z\in\mathcal{X}} \sqrt{\mathbb{P}(z)} \sum_{x\in\mathcal{X}} \frac{\mathbb{P}(x)}{\sqrt{\mathbb{P}(x)}} \\ &= \left[\sum_{x\in\mathcal{X}} \sqrt{\mathbb{P}(x)}\right]^2 = 2^{H_{1/2}(X)} \end{aligned}$$

which provides another *new operational definition of Rényi entropy of order* $1/2$ *relating it exactly to oblivious guessing.* If we compare this to Theorem 1, we can see that the penalty paid for oblivious memory constrained guessing, as opposed to using the optimal guessing sequence is roughly a factor of 2 in the expectation.

2.3 Power and Memory Constrained Guessor Minimizing the Probability of Failure

Consider the case where the guesses are still i.i.d., drawn from $\mathbb{Q}(x)$ but the guessor (maybe a sensor network node) decides *ahead of time* that she will only use $L \in \mathbb{N}$ guesses. Again we need to find the best $\mathbb{Q}(x)$, but now the appropriate criterion is minimizing the failure probability in L guesses, namely,

$$\mathbb{P}_{fail}(L) = \sum_{x \in \mathcal{X}} \mathbb{P}(x)(1 - \mathbb{Q}(x))^L$$

which yields the Lagrangian

$$\begin{aligned} J &= \mathbb{P}_{fail}(L) + \lambda(\sum_{x \in \mathcal{X}} \mathbb{Q}(x) - 1) \\ &= \sum_{x \in \mathcal{X}} \mathbb{P}(x)(1 - \mathbb{Q}(x))^L + \lambda(\sum_{x \in \mathcal{X}} \mathbb{Q}(x) - 1), \end{aligned}$$

which directly leads to the conditions

$$\frac{\partial J}{\partial \mathbb{Q}(x)} = -L\mathbb{P}(x)(1 - \mathbb{Q}(x))^{L-1} + \lambda = 0, \qquad \forall x \in \mathcal{X}$$

and since L is constant, we obtain after rearranging the equation above,

$$\mathbb{Q}(x) = 1 - \left(\frac{\mu}{\mathbb{P}(x)}\right)^{1/(L-1)}$$

for some positive constant $\mu = \lambda/L$. The second derivative is

$$\frac{\partial^2 J}{\partial \mathbb{Q}(x)^2} = L(L-1)\mathbb{P}(x)(1 - \mathbb{Q}(x))^{L-2}$$

and if we once again assume the non-degeneracy condition $0 < \mathbb{Q}(x) < 1$ for all $x \in \mathcal{X}$ we conclude that the second derivative is positive indicating that the stationary point determined minimizes the probability $\mathbb{P}_{fail}(L)$. Note that normalization requires that we have $\sum_{x \in \mathcal{X}} \mathbb{Q}(x) = 1$ which then yields

$$\sum_{x \in \mathcal{X}} \left[1 - \left(\frac{\mu}{\mathbb{P}(x)}\right)^{1/(L-1)}\right] = 1$$

or

$$\sum_{x\in\mathcal{X}}\left(\frac{\mu}{\mathbb{P}(x)}\right)^{1/(L-1)} = |\mathcal{X}| - 1$$

which can be satisfied by choosing μ as below,

$$\mu = \left(\frac{|\mathcal{X}| - 1}{\sum_{x\in\mathcal{X}} \mathbb{P}(x)^{-1/(L-1)}}\right)^{L-1},$$

thus proving Theorem 6.

Theorem 6. *If the attacker is restricted to L guesses, her optimal oblivious strategy is to generate L i.i.d. guesses from the following distribution*

$$\mathbb{Q}(x) = 1 - \left[\frac{|\mathcal{X}| - 1}{\sum_{z\in\mathcal{X}} \left(\mathbb{P}(x)/\mathbb{P}(z)\right)^{-1/(L-1)}}\right], \qquad \forall x \in \mathcal{X}$$

Note that in this case, a power sum related to the probability distribution $\mathbb{P}(x)$ is also involved, however this is not a Rényi entropy since the exponent $-1/(L-1)$ is negative. Neither is it related to any other kind of generalized entropy in the literature. However, if we rewrite it as below

$$\mathbb{Q}(x) = 1 - \mathbb{P}(x)^{-1/(L-1)}\left[\frac{|\mathcal{X}| - 1}{\sum_{x\in\mathcal{X}} \mathbb{P}(x)^{-1/(L-1)}}\right], \qquad \forall x \in \mathcal{X}$$

we note that it is related to the weighted power means that are widely used in mathematical analysis [13], namely

$$M_n^{[r]}(a, w) = \left(\frac{\sum_{k=1}^n w_k a_k^r}{\sum_{k=1}^n w_i}\right)^{1/r}$$

which are defined for real quantities $r \neq 0$ where $a = (a_1, \ldots, a_n)$ and $w = (w_1, \ldots, w_n)$. Thus we can write the above equation as

$$\mathbb{Q}(x) = 1 - \frac{\mathbb{P}(x)^{-1/(L-1)}(N-1)}{\left(M_N^{[-L/(L-1)]}(\mathbb{P}, \mathbb{P})\right)^{-L/(L-1)}}$$

where we define $\mathbb{P} = (\mathbb{P}(x_1), \ldots, \mathbb{P}(x_N))$ with
$\mathcal{X} = \{x_1, \ldots, x_N\}$. The power means themselves are also related to the Shannon entropy in the sense that if we let $(w_1, \ldots, w_n)$ and $(a_1, \ldots, a_n)$ be the same finite probability distribution $\mathbb{P}$ then we have

$$\lim_{r\to 0} \log M_N^{[r]}(\mathbb{P}, \mathbb{P}) = H(\mathbb{P}).$$

3 Multiple Guessors Case

We now want to consider the case of an attack that is more distributed, perhaps attacking multiple targets, whose passwords are assumed to come from the same distribution $\mathbb{P}(x)$. The question we want to answer is the following:

Given $\mathbb{P}(x)$, how should the attackers choose a distribution $\mathbb{Q}(x)$ in order to optimize some performance criterion, when all the attacker does (resp. attackers do) is to draw random sequential guesses from $\mathbb{Q}(x)$?

3.1 Multiple Memory Constrained Guessors Using Distributed Oblivious Guessing

We briefly discuss the case of $v \geq 2$ guessors working in parallel, each drawing from the same distribution $\mathbb{Q}(x)$, but not coordinating their guesses. First we note that if the guessors use some distribution $\mathbb{Q}(x)$ and they collectively work at a rate equal to v times the rate of the single guessor considered in the subsection above, their performance will be within the bounds

$$\left\lfloor \frac{\mathbb{E}_{\mathbb{Q}}[G]}{v} \right\rfloor \leq \mathbb{E}_{\mathbb{Q}}[G_v] \leq \left\lceil \frac{\mathbb{E}_{\mathbb{Q}}[G]}{v} \right\rceil$$

where we have used the notation $\mathbb{E}_{\mathbb{Q}}[G_v]$ for the expected number of guesses when v guessors each use $\mathbb{Q}(x)$. We now address the issue of optimizing the distribution $\mathbb{Q}(x)$ once v is fixed, instead of using the $\mathbb{Q}$ from the subsection above. From the rest of the subsection, we drop the subscript $\mathbb{Q}$ from the expectations for simplicity of presentation. Note that we can write

$$P[G_v = k] = Pr[G \in [(k-1)v + 1, kv] \cap \mathbb{Z}^+]$$

which gives

$$\mathbb{E}[G_v] = \sum_{x \in \mathcal{X}} \mathbb{P}(x)\mathbb{Q}(x) \sum_{k=0}^{\infty} (1+k)[(1-\mathbb{Q}(x))^v]^k \sum_{j=1}^{v} (1-\mathbb{Q}(x))^{j-1},$$

which can be simplified to

$$\mathbb{E}[G_v] =$$

$$\sum_{x\in\mathcal{X}}\mathbb{P}(x)\mathbb{Q}(x)\sum_{k=0}^{\infty}(1+k)[(1-\mathbb{Q}(x))^v]^k\left[\frac{1-(1-\mathbb{Q}(x))^v}{\mathbb{Q}(x)}\right],$$

if we use the sum of a finite geometric series. Using once again the generating function identity in (4) yields

$$\mathbb{E}[G_v] =$$

$$\sum_{x\in\mathcal{X}}\mathbb{P}(x)\mathbb{Q}(x)\left[1-(1-\mathbb{Q}(x))^v\right]^{-2}\left[\frac{1-(1-\mathbb{Q}(x))^v}{\mathbb{Q}(x)}\right]$$

or

$$\mathbb{E}[G_v]=\sum_{x\in\mathcal{X}}\left(\frac{\mathbb{P}(x)}{1-(1-\mathbb{Q}(x))^v}\right).$$

Now we define the Lagrangian

$$J_v=\mathbb{E}[G_v]+\lambda(\sum_{x\in\mathcal{X}}\mathbb{Q}(x)-1)$$

which upon differentiation gives

$$\frac{\partial J_v}{\partial\mathbb{Q}(x)}=-\frac{\mathbb{P}(x)v(1-\mathbb{Q}(x))^{v-1}}{(1-(1-\mathbb{Q}(x))^v)^2}+\lambda=0,\qquad\forall x\in\mathcal{X}$$

which indicates that the optimum distribution $\mathbb{Q}(x)$ satisfies

$$\frac{v(1-\mathbb{Q}(x))^{v-1}}{(1-(1-\mathbb{Q}(x))^v)^2}\propto\frac{1}{\mathbb{P}(x)}.$$

Let's define $R(x)=1-\mathbb{Q}(x)$ which takes on values in $(0,1)$ (note that R is not a probability distribution) which means that we have

$$\frac{(1-R(x)^v)^2}{vR(x)^{v-1}}\propto\mathbb{P}(x).$$

If we now consider the function

$$f(u)=\frac{(1-u^v)^2}{vu^{v-1}}\tag{5}$$

defined on $(0,1)$ we can see that its derivative can be simplified to

$$f'(u)=-\frac{2u(1-u^v)+(v-1)(1-u^v)^2}{vu^v}$$

which is clearly negative for integer $v \geq 2$ thus $f : (0, 1) \to \mathbb{R}$ is a monotone strictly decreasing one-to-one mapping and thus is invertible. We have thus proved

Theorem 7. *If v attackers are using distributed guessing their optimal oblivious strategy is to independently generate their guesses according to the distribution*

$$\mathbb{Q}(x) \ \propto \ \left[1 - f^{-1}(\mathbb{P}(x))\right]$$

which can then be normalized. Note that obtaining the optimal distribution here can be achieved by standard numerical methods.

4 Conclusions

The new results we obtained in this paper have provided a complement to the existing results, and have applications in key guessing attacks in three main scenarios.

1. Single, unrestricted attacker.
2. Single, constrained attacker.
3. Multiple distributed attackers.

In Theorems 2 and 3, we obtained a novel and explicit exact characterization of expected number of guesses for a single attacker in the unrestricted attacker case in terms of Renyi entropy. Another new result obtained in the paper is Theorem 4, an interesting derivation of the union bound which is used widely in information theory, in terms of the expected number of guesses in a conditional guessing scheme which takes the output of a communication channel as its input.

References

[1] E. Arikan; An Inequality on Guessing and Its Application to Sequential Decoding, *IEEE Transactions on Information Theory,* 42(1):99-105, 1996.

[2] E. Arikan and N. Merhav; Guessing subject to distortion, *IEEE Transactions on Information Theory,* 44(3):1041-1056, 1998.

[3] E. Arikan and N. Merhav; Joint Source-channel Coding and Guessing with Application to Sequential Decoding, *IEEE Transactions on Information Theory,* 44(5):1756-1769, 1998.

[4] S. Boztaş; Comments on 'An Inequality on Guessing and Its Application to Sequential Decoding', *IEEE Transactions Information Theory,* 43(6):2062-2063, 1997.

[5] S. Boztaş; Oblivious Distributed Guessing , *Proc. IEEE International Symposium on Information Theory,* 2161-2165, 2012.

[6] S.S. Dragomir and S. Boztaş; Some Estimates of the Average Number of Guesses to Determine a Random Variable, *Proc. IEEE International Symposium on Information Theory,* 1997.

[7] S.S. Dragomir and S. Boztaş; Estimation of Arithmetic Means and Their Applications in Guessing Theory, *Mathematical and Computer Modelling,* 28(10):31-43, 1998.

[8] J. L. Massey; Guessing and entropy, *Proc. 1994 IEEE International Symposium on Information Theory*, p. 204, 1994.

[9] D. Malone, W.G. Sullivan; Guesswork and entropy, *IEEE Transactions Information Theory,* 50(3):525- 526, 2004.

[10] M. Feder and N. Merhav; Relations between entropy and Error Probability, *IEEE Transactions on Information Theory* 40(1):259-266, 1994.

[11] N. Merhav and E. Arikan; The Shannon Cipher System with a Guessing Wiretapper, it IEEE Transactions on Information Theory, 45(6):1860-1866, 1999.

[12] N. Merhav, R.M. Roth, E. Arikan; Hierarchical guessing with a fidelity criterion, *IEEE Transactions Information Theory,* 45(1):330-337, 1999.

[13] D.S. Mitrinović; *Analytic Inequalities*, Springer,1970.

[14] C.-E. Pfister, W.G. Sullivan; Rényi Entropy, Guesswork Moments, and Large Deviations, *IEEE Transactions on Information Theory*, 50(11):2794, 2004.

[15] J. O. Pliam; On the incomparability of Entropy and Marginal Guesswork in Brute-force Attacks, Proc. INDOCRYPT 2000, *Lecture Notes in Computer Science* 1977:67–79, 2000.

[16] R. Sundaresan; Guessing Under Source Uncertainty, *IEEE Transactions on Information Theory* 53(1): 269 - 287, 2007.

[17] M. K. Hanawal and R. Sundaresan; Randomised Attacks on Passwords, *Technical Report TR-PME-2010-11*, Dept. ECE, Indian Institute of Science, Bangalore, available at
http://www.pal.ece.iisc.ernet.in/PAM/docs/techreports/tech_rep10/

[18] R. Sundaresan; Guessing and Compression Subject to Distortion, *Technical Report TR-PME-2010-12*, Dept. ECE, Indian Institute of Science, Bangalore, available at
http://www.pal.ece.iisc.ernet.in/PAM/docs/techreports/tech_rep10/