

Guesswork Subject to a Total Entropy Budget

Arman Rezaee, Ahmad Beirami, Ali Makhdoumi, Muriel Médard, and Ken Duffy

Abstract

We consider an abstraction of computational security in password protected systems where a user draws a secret string of given length with i.i.d. characters from a finite alphabet, and an adversary would like to identify the secret string by querying, or guessing, the identity of the string. The concept of a “total entropy budget” on the chosen word by the user is natural, otherwise the chosen password would have arbitrary length and complexity. One intuitively expects that a password chosen from the uniform distribution is more secure. This is not the case, however, if we are considering only the average guesswork of the adversary when the user is subject to a total entropy budget. The optimality of the uniform distribution for the user’s secret string holds when we have also a budget on the guessing adversary. We suppose that the user is subject to a “total entropy budget” for choosing the secret string, whereas the computational capability of the adversary is determined by his “total guesswork budget.” We study the regime where the adversary’s chances are exponentially small in guessing the secret string chosen subject to a total entropy budget. We introduce a certain notion of uniformity and show that a more uniform source will provide better protection against the adversary in terms of his chances of success in guessing the secret string. In contrast, the average number of queries that it takes the adversary to identify the secret string is smaller for the more uniform secret string subject to the same total entropy budget.

I. INTRODUCTION

We consider the problem of identifying the realization of a discrete random variable X by repeatedly asking questions of the form: “Is x the identity of X ?”. This problem has been extensively studied by cryptanalysts who try to identify a secret key by exhaustively trying out all possible keys, where it is usually assumed that the secret key is drawn uniformly at random. We consider an n -tuple $X^n :=$

A. Rezaee, A. Beirami, A. Makhdoumi, and M. Médard are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. Emails: {armanr, beirami, makhdoum, medard}@mit.edu.

K. Duffy is with the National University of Ireland Maynooth, Ireland. Email: ken.duffy@nuim.ie.

X_1, \dots, X_n drawn from an i.i.d. source, $\mu_\theta(\cdot)$ on a finite alphabet \mathcal{X} , where θ represents the corresponding categorical distribution, which is not necessarily uniform. We measure security against a brute-force attacker who knows the source statistics completely, and who would query all the secret strings one by one until he is successful.

Denoting the number of guesses by $G_\theta^n(X^n)$, the optimal strategy of the attacker that minimizes the expected number of queries $\mathbb{E}[G_\theta^n(X^n)]$ is to guess the possible realizations of X^n in order of decreasing probability under $\mu_\theta^n(\cdot)$. Massey [1] proved that the Shannon entropy of X^n , $H(X^n)$, is a lower bound on the rate of growth of the expected guesswork, yet there is no upper bound on $\mathbb{E}[G_\theta^n(X^n)]$ in terms of $H(X^n)$. Arkan [2] proved that when we consider a string of growing length whose characters are drawn i.i.d, the positive moments of guesswork associated with the optimal strategy grow exponentially, and the exponents are related to the Rényi entropies of the single letter distribution:¹

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta [(G_\theta^n(X^n))^\rho] = H_{1/(1+\rho)}(X), \quad (1)$$

where the Rényi entropy of order ρ is

$$H_\rho(X) = \frac{1}{1-\rho} \log \left(\sum_{x \in \mathcal{X}} P(X=x)^\rho \right). \quad (2)$$

Note that $\lim_{\rho \rightarrow 0} H_\rho(X) = H(X)$ recovers the Shannon entropy. We also use the notations $H_\rho(\theta)$ and $H_\rho(X)$ interchangeably to refer to the Rényi entropy of a string drawn from a source with parameter vector θ . Although these connections have been extended to more general stochastic processes [3], [4], in this paper, we focus on i.i.d. processes for the sake of clarity of presentation.

Christiansen and Duffy [5] showed that the sequence $\{n^{-1} \log G_\theta^n(X^n)\}$ satisfies a Large Deviations Principle (LDP) and characterized its rate function, Λ_θ^* . Beirami *et al.* [6], [7] showed that Λ_θ^* can be expressed as a parametric function of the value of a “tilt” in a family of tilted distributions.

We remark that when the metric of difficulty is the growth rate in the expected number of guesses as a function of string length, the challenge for the adversary remains the same even if the adversary does not know the source statistics [8], [9].

In this paper, we first show a counter intuitive result that the average guesswork increases when the source becomes “less uniform” if the user is subject to a total entropy budget on the secret string. Next, we introduce a natural notion of total guesswork budget on the attacker and show that the probability of success of an adversary subject to a total guesswork budget increases when the source becomes “less uniform,” which is consistent with our intuition of choosing uniform passwords. We will formalize these notions in the rest of this paper.

¹In this paper, $\log(\cdot)$ denotes the natural logarithm.

II. PROBLEM SETUP

Given a finite alphabet \mathcal{X} , a memoryless (i.i.d) source on \mathcal{X} is defined by the set of probabilities $\theta_i = P[X = x_i]$ for all $i \in [|\mathcal{X}|]$, where $[n] := \{1, \dots, n\}$ and $\sum_{i \in [|\mathcal{X}|]} \theta_i = 1$. Hence, θ is an element of the $(|\mathcal{X}|-1)$ -dimensional probability simplex. We define $\Theta_{|\mathcal{X}|}$ as the open set of all probability vectors θ such that $\theta_i > 0$ for all $i \in \{1, \dots, |\mathcal{X}|\}$, which also excludes the uniform source $u_{|\mathcal{X}|} = (1/|\mathcal{X}|, \dots, 1/|\mathcal{X}|)$.

The tilt operation plays a central role in the analysis, and is the basis for many of our derivations:

Definition 1 (tilted θ of order α [6]). For any $\alpha \in \mathbb{R}$, define $\tau(\theta, \alpha)$ as the “tilted θ of order α ”, where $\tau(\theta, \alpha) = (\tau_1(\theta, \alpha), \dots, \tau_{|\mathcal{X}|}(\theta, \alpha))$, where $\tau_i : \Theta_{|\mathcal{X}|} \times \mathbb{R} \rightarrow \Theta_{|\mathcal{X}|}$ for all $i \in [|\mathcal{X}|]$ is given by

$$\tau_i(\theta, \alpha) := \frac{\theta_i^\alpha}{\sum_{i=1}^{|\mathcal{X}|} \theta_i^\alpha}. \quad (3)$$

Definition 2 (tilted family of θ). Let $\Gamma_\theta^+ \in \Theta_{|\mathcal{X}|}$ denote the “tilted family of θ ” and be given by

$$\Gamma_\theta^+ := \{\tau(\theta, \alpha) : \alpha \in \mathbb{R}_{>0}\}. \quad (4)$$

Observe that $\Gamma_\theta^+ \in \Theta_{|\mathcal{X}|}$ is a continuum of stochastic vectors in the probability simplex. Thus, the tilted family of a memoryless string-source with parameter vector θ is comprised of a set of memoryless string-sources whose parameter vectors belong to the tilted family of the vector θ , i.e., Γ_θ^+ .

Definition 3 (high-entropy/low-entropy members of tilted family of θ). Let $\bar{\Gamma}_\theta^+$ and $\underline{\Gamma}_\theta^+$ denote the sets of high-entropy and low-entropy members of the tilted family of θ , respectively, and be given by:

$$\bar{\Gamma}_\theta^+ = \{\tau(\theta, \alpha)\}_{0 \leq \alpha < 1}, \quad \underline{\Gamma}_\theta^+ = \{\tau(\theta, \alpha)\}_{\alpha > 1}. \quad (5)$$

Hence, $\Gamma_\theta^+ = \bar{\Gamma}_\theta^+ \cup \underline{\Gamma}_\theta^+ \cup \theta$.

Figure 1 depicts the probability simplex of all possible ternary parameter vectors, $|\mathcal{X}| = 3$. The yellow star represents the distribution $\theta = (0.1, 0.2, 0.7)$. Note that the tilted family of θ is parametrized by α . At $\alpha = 0$, we get the uniform distribution $\tau(\theta, 0) = u_3 = (1/3, 1/3, 1/3)$ and as $\alpha \rightarrow \infty$, we get to the degenerate case of $(0, 0, 1)$. The high-entropy and low-entropy members of the tilted family of θ are represented by blue and red, respectively. Note that all distributions in the high-entropy set, $\bar{\Gamma}_\theta^+$, have Shannon entropies higher than that of θ and are closer to the uniform distribution in the KL divergence sense [7]. Hence, the higher entropy members of the tilted family are “more uniform” than the lower entropy members of the tilted family.

Definition 4 (entropy budget per source character). Let $h \in (0, \log |\mathcal{X}|]$ denote the entropy budget per source character such that the user is required to choose a secret string from an i.i.d. process with parameter vector θ with $H(\theta) = h$.

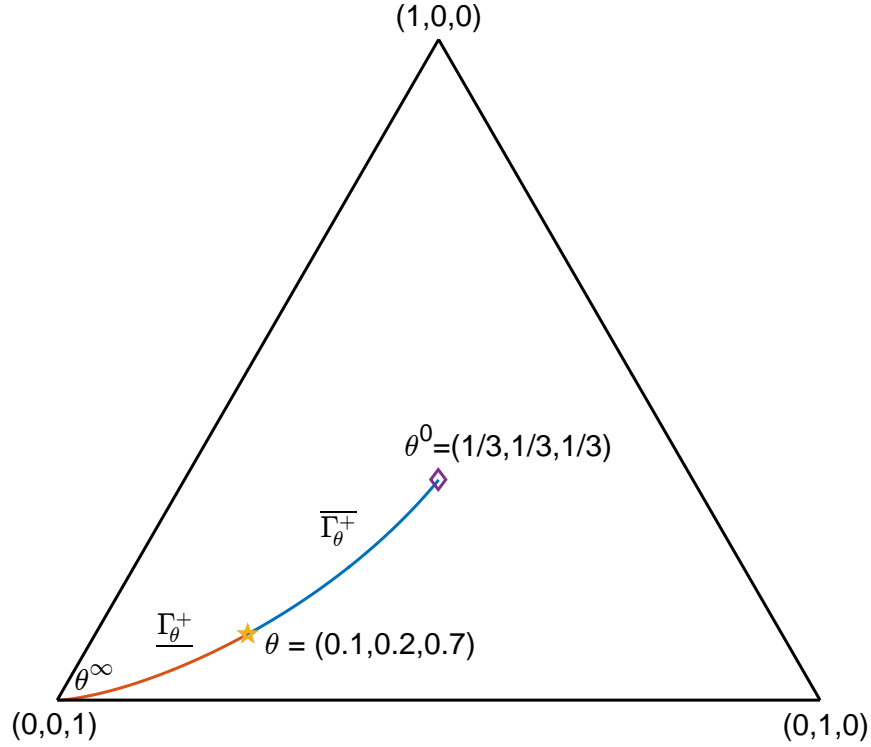


Figure 1: The probability simplex for a ternary alphabet. The figure represents the tilted family of $\theta = (0.1, 0.2, 0.7)$, as well as the high-entropy and low-entropy members of the family.

The concept of a total entropy budget on the entire secret string is a natural one or the user would choose an arbitrarily complex secret string. We use the entropy budget per source character defined above to ensure that the user is subject to the same total entropy budget by adjusting the length of the secret string for a fair comparison between string sources that have different entropy rates.

III. POSITIVE MOMENTS OF GUESSWORK

We first consider choosing strings with the same total (Shannon) entropy budget and measure security in terms of the positive moments of guesswork. If two sources have different entropy rates, we adjust the comparison by drawing a longer string from the lower entropy source. Formally, let us consider two sources with parameter vectors θ_1 and θ_2 on alphabet \mathcal{X} . Further, let $H(\theta_1)$ and $H(\theta_2)$ be the entropy rates of the two sources. Let the entropy ratio be

$$\eta := \frac{H(\theta_2)}{H(\theta_1)}. \quad (6)$$

Without loss of generality, throughout this paper we assume that $H(\theta_2) < H(\theta_1)$, and hence $0 < \eta < 1$. The user is given the option to choose a secret string from either of the two sources. For a fair comparison,

we assume that the entropy of the two strings is the same, $n_1 H(\theta_1) = n_2 H(\theta_2)$. That is

$$n_2 = \frac{1}{\eta} n_1. \quad (7)$$

To compare the growth rates of the positive moments of guesswork, in light of (1), we compare $H_{1/(1+\rho)}(\theta_1)$ and $\frac{1}{\eta} H_{1/(1+\rho)}(\theta_2)$. This will in turn impose the same total entropy budget on the strings drawn from the sources with parameter vectors θ_1 and θ_2 .

For a parameter vector θ , let an information random variable be defined as one that it takes the value $\log \frac{1}{\theta_i}$ with probability θ_i for all $i \in [|\mathcal{X}|]$. We need one more definition before we can state the result of this section:

Definition 5 (skewentropy condition (SEC)). A source with parameter vector $\theta \in \Theta_{|\mathcal{X}|}$ is said to satisfy the skewentropy condition (SEC) if

$$V(\theta)^2 + 2H(\theta)V(\theta) - H(\theta)S(\theta) > 0, \quad (8)$$

where $V(\theta)$ is the varentropy defined as the variance of an information random variable corresponding to θ :

$$V(\theta) := \sum_{i \in [|\mathcal{X}|]} \theta_i \left(\log \frac{1}{\theta_i} - H(\theta) \right)^2. \quad (9)$$

and $S(\theta)$ is the skewentropy, which is the skewness of an information random variable corresponding to θ :

$$S(\theta) := \sum_{i \in [|\mathcal{X}|]} \theta_i \left(\log \frac{1}{\theta_i} - H(\theta) \right)^3. \quad (10)$$

Note that varentropy has been studied extensively and naturally arises in the finite block length information theory [10], [11], and more recently in the study of polar codes [12]. To the best of our knowledge, skewentropy has not been studied before, and we provide some properties of the SEC in Section V.

Equipped with this definition, we provide an ordering of the sources that belong to the same tilted family.

Theorem 1. Let $\theta_1 \in \Theta_{|\mathcal{X}|}$. For any $\theta_2 \in \Gamma_{\theta_1}^+$,

$$H_{1/(1+\rho)}(\theta_1) < \frac{1}{\eta} H_{1/(1+\rho)}(\theta_2) \quad \forall \rho > 0, \quad (11)$$

if and only if θ_1 satisfies the SEC in Definition 5. Note that η is the entropy ratio defined in (6).

The proof is provided in the appendix. Theorem 1 provides a natural ordering of sources that belong to the same tilted family. The “less uniform” low per-character entropy members of the tilted family take

exponentially more number of queries, on the average, to breach compared to their more uniform higher per character entropy counterparts.

Corollary 2. Let $u_{|\mathcal{X}|}$ denote the uniform source. Then for any $\theta \in \Theta_{|\mathcal{X}|}$, and any $\rho > 0$,

$$\log |\mathcal{X}| = H_{1/(1+\rho)}(u_{|\mathcal{X}|}) < \frac{1}{\eta} H_{1/(1+\rho)}(\theta),$$

where $\eta = H(\theta)/\log |\mathcal{X}|$.

Corollary 2 suggests that, of all sources whose parameter vectors are in the (interior of the) probability simplex, the uniform source is the easiest to breach in terms of the positive moments of guesswork when the user is subject to a total entropy budget. This is in contrast to our intuition that more uniformity provides better security.

IV. PROBABILITY OF SUCCESS SUBJECT TO A GUESSWORK BUDGET

In this section, we put forth a natural notion of total guesswork budget, leading to a security metric consistent with our intuition. Similar to the case of an entropy budget, we need to define guesswork budget per source character for our analysis.

Definition 6 (guesswork budget per source character). Let $g \in (0, \log |\mathcal{X}|]$ denote the guesswork budget per source character, such that e^{gn} is the total number of queries that the inquisitor can make in order to identify a secret string of length n .

Note that by this definition, the inquisitor is supposed to possess the resources for querying an exponentially growing number of strings (with the sequence length). In particular, $g = \log |\mathcal{X}|$ corresponds to an adversary who is capable of querying all of the possible $|\mathcal{X}|^n$ outcomes of the source to successfully identify the secret string with probability 1.

Lemma 1. If $g < H(\theta)$, then

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta[G_\theta(X^n) \leq e^{gn}] = 0,$$

and if $g > H(\theta)$, then

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta[G_\theta(X^n) \leq e^{gn}] = 1.$$

Recall that Arikan [2] showed that the growth rate of the moments of guesswork is governed by atypical sequences resulting in the appearance of the Rényi entropies in the expression. On the other hand, Lemma 1 states that the cutoff for the adversary to be successful with high probability is still governed by the Shannon entropy (as intuitively expected).

In the regime where $g < H(\theta)$, we would like to study the behavior of correct guessing. The next lemma relates the exponent of an exponentially large number of possible guesses to the LDP rate function.

Lemma 2. If $g < H(\theta)$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\mathbb{P}_\theta[G_\theta(X^n) \leq e^{gn}] } = \Lambda_\theta^*(g). \quad (12)$$

Hence, $\mathbb{P}_\theta[G_\theta(X^n) \leq e^{gn}] \approx e^{-n\Lambda_\theta^*(g)}$, and a larger $\Lambda_\theta^*(g)$ directly implies a more secure source against a brute-force attacker who is subject to a guesswork budget g for a fixed n . We use the above rate function as the metric for comparing two string-sources given a total guesswork budget, naturally defined as $g \times n$.

Using the notion of the tilt, we can represent the rate function $\Lambda_\theta^*(g)$ as a parametric function of α for a family of tilted distributions. The rate function, $\Lambda_\theta^*(g)$, associated with $\theta \in \Theta_{|\mathcal{X}|}$ can be directly computed as [7]:

$$\Lambda_\theta^*(g) = D(\tau(\theta, \alpha(g)) \parallel \theta), \quad (13)$$

for $\alpha(g) = \arg_{\alpha \in \mathbb{R}^+} \{H(\tau(\theta, \alpha)) = g\}$. This characterization plays a central role in our derivations.

Recall that we adjust the string lengths in order to make sure that the secret string chosen by the user is subject to a given total entropy budget. As the idea of the total guesswork budget is that the adversary can make a fixed number of queries regardless of the source from which the user is choosing the password, we compare the sources in terms of the probability of success subject to an adjusted guesswork budget per source character (see (12)). To keep the total guessing budget of the adversary the same, i.e., $e^{n_1 g_1} = e^{n_2 g_2}$, we must adjust the guesswork budget per source character as follows:

$$g_2 = \eta g_1. \quad (14)$$

In light of (14), we compare $\Lambda_{\theta_1}^*(g_1)$ with $\frac{1}{\eta} \Lambda_{\theta_2}^*(g_2) = \frac{1}{\eta} \Lambda_{\theta_2}^*(\eta g_1)$ for sources with parameter vectors θ_1 and θ_2 .

We are now ready to provide our results on the adversary's probability of success.

Theorem 3. Let $\theta_1 \in \Theta_{|\mathcal{X}|}$. For any $\theta_2 \in \underline{\Gamma}_{\theta_1}^+$,

$$\Lambda_{\theta_1}^*(g_1) > \frac{1}{\eta} \Lambda_{\theta_2}^*(g_2), \quad \forall g_1 < H(\theta_1), \quad (15)$$

if and only if θ_1 satisfies the SEC (see Definition 5).

We remark that the same SEC appears to be the crucial quantity for the statement of Theorem 3 to hold. This theorem implies that when the adversary is subject to a guesswork budget g_1 (i.e., he can only submit $e^{n_1 g_1}$ queries to identify a secret string of length n) for some $g_1 \in (0, H(\theta_1))$, then the chances of

correctly identifying the random string produced by a “more uniform” high per-character entropy member of the tilted family is exponentially smaller than that of the less uniform low per-character entropy source belonging to the same tilted family so long as the source satisfies the SEC when the user is subject to the same total entropy budget and the adversary is subject to the same total guesswork budget. In particular, the uniform source is the most secure against such an adversary subject to a guesswork budget:

Corollary 4. Let $u_{|\mathcal{X}|}$ denote the uniform information source. Then, for any $\theta \in \Theta_{|\mathcal{X}|}$ and $g < \log |\mathcal{X}|$, we have

$$\log |\mathcal{X}| - g = \Lambda_{u_{|\mathcal{X}|}}^*(g) > \frac{1}{\eta} \Lambda_{\theta}^*(\eta g), \quad (16)$$

where $\eta = H(\theta)/\log |\mathcal{X}|$.

We remark that these security guarantees are against an adversary that is not powerful enough to be able to explore the entire typical set rendering his chances of success exponentially small. The “more uniform” sources provide an exponentially smaller chance to such an adversary to be successful.

We emphasize that the implications of Theorems 1 and 3 are in stark contrast to each other. On the one hand, more uniformity results in an exponential decrease in the number of queries expected of an adversary to correctly identify a secret string when the user is subject to a total entropy budget (Theorem 1). On the other hand, more uniformity decreases the chances of an adversary in identifying the secret string when the adversary’s power is limited by a total guesswork budget as well (Theorem 3).

V. PROPERTIES OF THE SEC

Noting that SEC introduced in Definition 5 is a new concept, we study this condition in more detail in this section. Let us start with the binary memoryless sources.

Lemma 3. Let $\theta \in \Theta_2$. Further, let $\phi = \min\{\theta_1, \theta_2\} < \frac{1}{2}$. Then,

$$H(\theta) = \phi \log \left(\frac{1}{\phi} \right) + (1 - \phi) \log \left(\frac{1}{1 - \phi} \right), \quad (17)$$

$$V(\theta) = \phi(1 - \phi) \log^2 \left(\frac{1 - \phi}{\phi} \right), \quad (18)$$

$$S(\theta) = \phi(1 - \phi)(1 - 2\phi) \log^3 \left(\frac{1 - \phi}{\phi} \right). \quad (19)$$

The next theorem is our main result for binary memoryless sources:

Theorem 5. Any $\theta \in \Theta_2$ satisfies the SEC.

While Theorem 5 shows that all binary memoryless sources satisfy the SEC, the same argument does not extend to larger alphabets.

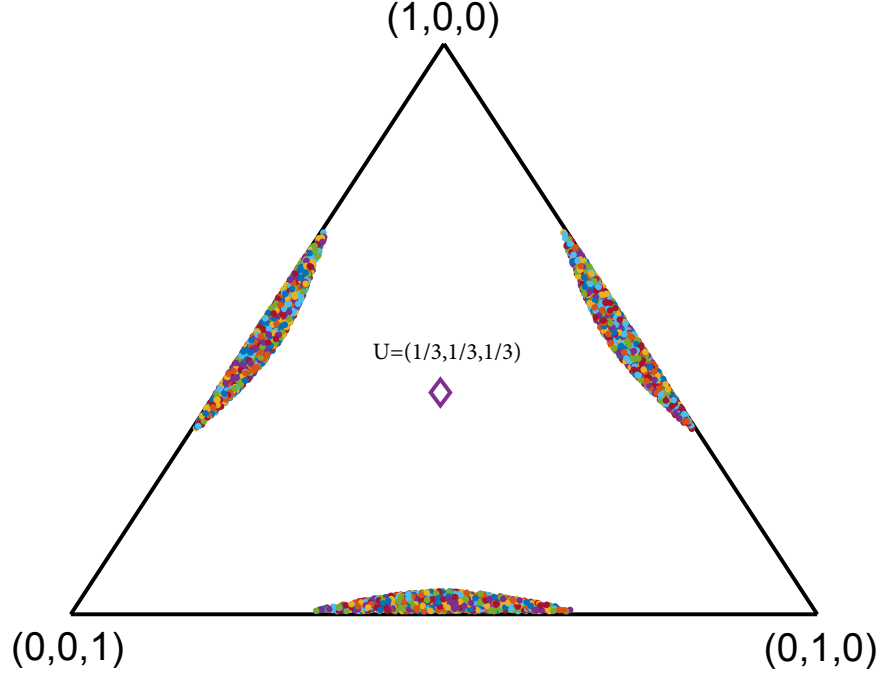


Figure 2: Depiction of the probability simplex for a ternary alphabet. The figure represents the set of distributions that do not satisfy the SEC.

Theorem 6. For any $|\mathcal{X}| > 2$, there exists $\theta \in \Theta_{|\mathcal{X}|}$, such that θ does not satisfy the SEC.

Despite the negative result in Theorem 6, we show that sources that are approximately uniform satisfy the SEC for any alphabet size. Here is the key result for such sources:

Theorem 7. Suppose that $\theta \in \Theta_{|\mathcal{X}|}$ is such that

$$\left| \log \frac{1}{\theta_i} - H(\theta) \right| < 2, \quad \forall i \in [|\mathcal{X}|]. \quad (20)$$

Then θ satisfies the SEC.

As a corollary, we state the condition more explicitly in terms of θ_i 's.

Corollary 8. Suppose that $\theta \in \Theta_{|\mathcal{X}|}$ is such that

$$\frac{e^{-1}}{|\mathcal{X}|} < \theta_i < \frac{e}{|\mathcal{X}|}, \quad \forall i \in [|\mathcal{X}|]. \quad (21)$$

Then, θ satisfies the SEC.

Figure 2 depicts the set of ternary distributions that do not satisfy the SEC. As can be seen, source close to uniform satisfy the SEC while sources that are close to uniform on a two-dimensional alphabet while almost missing the third character in the alphabet do not satisfy the SEC.

VI. NUMERICAL EXPERIMENTS

In this section, we provide some numerical experiments. We compare several binary sources, where $\theta = (\theta_1, \theta_2)$ is the source parameter vector. The parameter vectors used for the experiments are listed in Table I. The length and the parameter vector are chosen such that $nH(\theta) = 9 \log 2$ nats for all of the pairs. Although the theorems proved in this paper are of asymptotic nature, we have chosen to run experiments on finite-length sequences instead to emphasize the applicability of the results even in very short lengths. As can be seen in Fig. 3, as the entropy rate of the source decreases, the moments of guesswork increase exponentially subject to the same entropy budget. On the other hand, as shown in Fig. 4, as the entropy rate of the source decreases, the chances of an adversary subject to a fixed total guesswork budget increases, which is consistent with our intuition.

θ_1	n
0.5000	9
0.3160	10
0.2145	12
0.1461	15
0.1100	18
0.0820	22

Table I: The list of source parameters and sequence lengths of binary sources used in the experiments.

VII. CONCLUSION

In this paper, we studied guesswork subject to a total entropy budget. We showed that the conclusions about security deduced from the analysis of the average guesswork could be counter-intuitive in that they suggest that the uniform source is not the strongest source against brute-force attacks. To remedy the problem, we introduced the concept of total guesswork budget, and showed that if the adversary is subject to a total guesswork budget, the uniform source provides the strongest security guarantees against the brute-force attacker, which is consistent with our intuition.

APPENDIX

PROOFS

Proof of Theorem 1: This is equivalent to showing that for all $\rho > 0$,

$$\frac{H_{1/(1+\rho)}(\theta_2)}{H(\theta_2)} > \frac{H_{1/(1+\rho)}(\theta_1)}{H(\theta_1)} \quad (22)$$

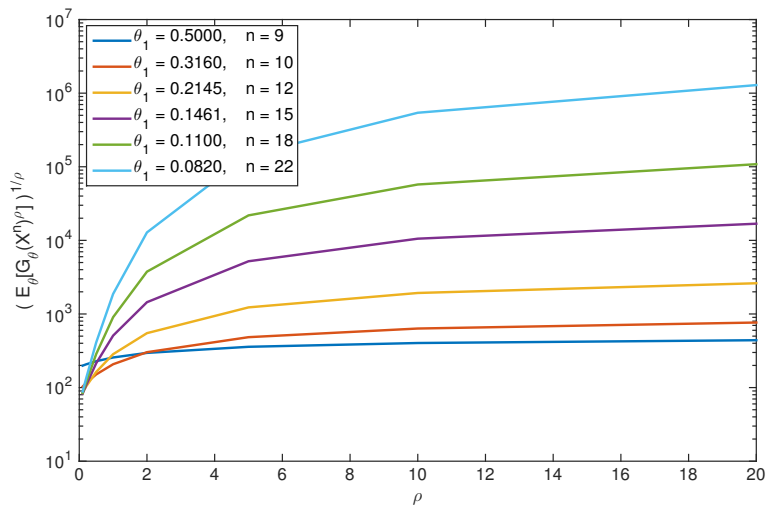


Figure 3: The positive moments of guesswork for sources subject to the same total entropy budget in Table I.

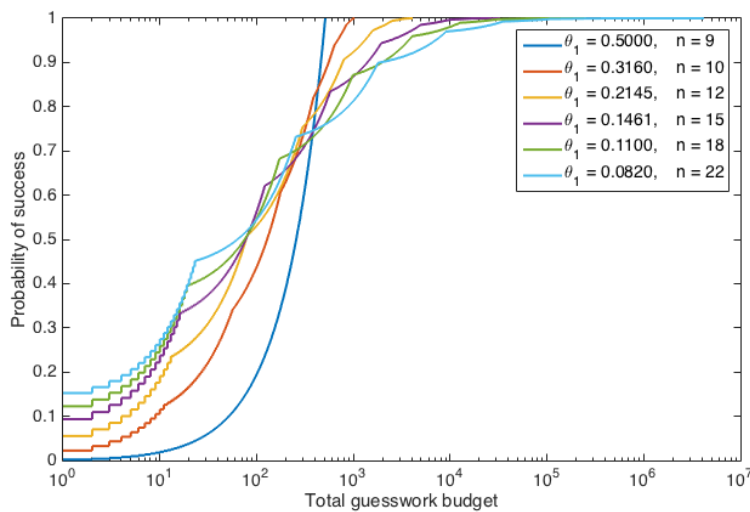


Figure 4: The probability of success as a function of the total guesswork budget for binary sources of Table I subject to the same total entropy budget.

for all $\theta_2 \in \underline{\Gamma}_\theta^+$. Let $\beta := 1/(1 + \rho)$, and hence $\beta < 1$. The statement above is in turn equivalent to showing:

$$\frac{\partial}{\partial \alpha} \left[\frac{H_\beta(\tau(\theta_1, \alpha))}{H(\tau(\theta_1, \alpha))} \right]_{\alpha=1} > 0, \quad \forall \beta < 1. \quad (23)$$

It is straightforward to show that (76) is equivalent to

$$\frac{\frac{\partial}{\partial \alpha} [H_\beta(\tau(\theta_1, \alpha))]_{\alpha=1}}{H_\beta(\theta_1)} > \frac{\frac{\partial}{\partial \alpha} [H(\tau(\theta_1, \alpha))]_{\alpha=1}}{H(\theta_1)}, \quad \forall \beta < 1. \quad (24)$$

Finally, we prove the following statement that is equivalent to (24):

$$\frac{\partial}{\partial \beta} \left[\frac{\frac{\partial}{\partial \alpha} [H_\beta(\tau(\theta_1, \alpha))]_{\alpha=1}}{H_\beta(\theta_1)} \right]_{\beta=1} < 0. \quad (25)$$

This is equivalent to showing:

$$\begin{aligned} & \frac{\partial^2}{\partial \alpha \partial \beta} [H_\beta(\tau(\theta_1, \alpha))]_{\alpha=\beta=1} \quad H(\theta_1) \\ & < \frac{\partial}{\partial \beta} [H_\beta(\theta_1)]_{\beta=1} \quad \frac{\partial}{\partial \alpha} [H(\tau(\theta_1, \alpha))]_{\alpha=1}. \end{aligned} \quad (26)$$

The above statement is shown to hold if and only if θ_1 satisfies the SEC (Definition 5) invoking Lemmas 4, 5, and 6, which completes the proof of the theorem. \blacksquare

Lemma 4. For all $\theta \in \Theta_{|\mathcal{X}|}$, we have

$$\frac{\partial}{\partial \alpha} [H(\tau(\theta, \alpha))]_{\alpha=1} = -V(\theta). \quad (27)$$

See [7] for the proof.

Lemma 5. For all $\theta \in \Theta_{|\mathcal{X}|}$, we have

$$\frac{\partial}{\partial \beta} [H_\beta(\theta)]_{\beta=1} = -\frac{1}{2}V(\theta). \quad (28)$$

See [7] for the proof.

Lemma 6. For all $\theta \in \Theta_{|\mathcal{X}|}$, we have

$$\frac{\partial^2}{\partial \alpha \partial \beta} [H_\beta(\tau(\theta, \alpha))]_{\alpha=\beta=1} = -V(\theta) + \frac{1}{2}S(\theta). \quad (29)$$

Proof: It is proved in [7] that

$$\frac{\partial}{\partial \alpha} [H_\beta(\tau(\theta, \alpha))]_{\alpha=1} = \frac{\beta}{1-\beta} (H(\theta) - H(\tau(\theta, \beta)|\theta)). \quad (30)$$

Hence, we differentiate with respect to β to get:

$$\begin{aligned} \frac{\partial^2}{\partial \alpha \partial \beta} [H_\beta(\tau(\theta, \alpha))]_{\alpha=1} &= \frac{1}{(1-\beta)^2} (H(\theta) - H(\tau(\theta, \beta)|\theta)) \\ &+ \frac{\beta}{1-\beta} V(\tau(\theta, \beta)|\theta). \end{aligned}$$

Next, we take the limit as $\beta \rightarrow 1$, and by applying L'Hospital's rule we arrive at:

$$\frac{\partial^2}{\partial \alpha \partial \beta} [H_\beta(\tau(\theta, \alpha))]_{\alpha=\beta=1} = -V(\theta) - \frac{1}{2} \frac{\partial}{\partial \beta} [V(\tau(\theta, \beta) || \theta)]_{\beta=1}. \quad (31)$$

Finally, the proof is completed by invoking Lemma 7. ■

Lemma 7. For any $\theta \in \Theta_{|\mathcal{X}|}$,

$$\frac{\partial}{\partial \alpha} [V(\tau(\theta, \alpha) || \theta)]_{\alpha=1} = -S(\theta),$$

where $S(\theta)$ is defined in (10).

Proof: By definition

$$\begin{aligned} & \left. \frac{\partial}{\partial \alpha} V(\tau(\theta, \alpha) || \theta) \right|_{\alpha=1} \\ &= \sum_{i \in [|\mathcal{X}|]} \left. \frac{\partial}{\partial \alpha} \tau_i(\theta, \alpha) \right|_{\alpha=1} \left(H(\tau(\theta, \alpha) || \theta) - \log \frac{1}{\theta_i} \right)^2 \\ &= \sum_{i \in [|\mathcal{X}|]} \theta_i \left(H(\tau(\theta, \alpha) || \theta) - \log \frac{1}{\theta_i} \right)^3 \end{aligned} \quad (32)$$

$$= -S(\theta), \quad (33)$$

where (32) follows by invoking Lemma 8 of [7]. ■

Proof of Theorem 3: Let us recall that $\theta_2 = \tau(\theta_1, \alpha)$ for some $\alpha > 1$. We can find t_1 and t_2 in the domain of each rate function such that the derivatives of the rate function are both equal to a constant $\rho > -1$. It follows from [2] that:

$$\begin{aligned} t_1 &= \arg_t \left\{ \frac{\partial}{\partial t} \Lambda_{\theta_1}^*(t) = \rho \right\} \Rightarrow t_1 = H(\tau(\theta_1, \beta)), \\ t_2 &= \arg_t \left\{ \frac{1}{\eta} \frac{\partial}{\partial t} \Lambda_{\theta_2}^*(\eta t) = \rho \right\} \Rightarrow t_2 = \frac{1}{\eta} H(\tau(\theta_2, \beta)), \end{aligned} \quad (34)$$

where $\beta = 1/(1 + \rho)$. We focus on $\rho < 0$, and hence $\beta \in (1, \infty)$. Note that $\beta = 1$, (equivalently $\rho = 0$) corresponds to the coinciding zeros of both rate functions. Once again recalling that the rate functions are convex, proving $(1/\eta) \Lambda_{\theta_2}^*(\eta t) > \Lambda_{\theta_1}^*(t)$ is equivalent to showing that $t_2 < t_1$ (as defined in (34)) for all $\beta > 1$. This is in turn equivalent to showing:

$$\frac{H(\tau(\theta_2, \beta))}{H(\theta_2)} < \frac{H(\tau(\theta_1, \beta))}{H(\theta_1)}, \quad \forall \alpha, \beta > 1. \quad (35)$$

This is equivalent to:

$$\frac{\partial}{\partial \alpha} \left[\frac{H(\tau(\theta_1, \alpha\beta))}{H(\tau(\theta_1, \alpha))} \right]_{\alpha=1} < 0, \quad \forall \beta > 1. \quad (36)$$

It is straightforward to show that (36) is equivalent to

$$\frac{\frac{\partial}{\partial \alpha} [H(\tau(\theta_1, \alpha\beta))]_{\alpha=1}}{H(\tau(\theta_1, \beta))} > \frac{\frac{\partial}{\partial \alpha} [H(\tau(\theta_1, \alpha))]_{\alpha=1}}{H(\theta_1)}, \quad \forall \beta > 1. \quad (37)$$

Finally, we prove the following statement that is equivalent to (37):

$$\frac{\partial}{\partial \beta} \left[\frac{\frac{\partial}{\partial \alpha} [H(\tau(\theta_1, \alpha\beta))]_{\alpha=1}}{H(\tau(\theta_1, \beta))} \right]_{\beta=1} < 0. \quad (38)$$

This is equivalent to showing:

$$\begin{aligned} & \frac{\partial^2}{\partial \alpha \partial \beta} [H(\tau(\theta_1, \alpha\beta))]_{\alpha=\beta=1} \quad H(\theta_1) \\ & < \frac{\partial}{\partial \beta} [H(\tau(\theta_1, \beta))]_{\beta=1} \quad \frac{\partial}{\partial \alpha} [H(\tau(\theta_1, \alpha))]_{\alpha=1}. \end{aligned} \quad (39)$$

The above statement is shown to hold if and only if θ_1 satisfies the SEC (Definition 5) invoking Lemmas 4 and 8, which completes the proof of the theorem. ■

Lemma 8. For all $\theta \in \Theta_{|\mathcal{X}|}$, we have

$$\frac{\partial^2}{\partial \alpha \partial \beta} [H(\tau(\theta, \alpha\beta))]_{\alpha=\beta=1} = -2V(\theta) + S(\theta). \quad (40)$$

Proof: Noting that $\tau(\theta, \alpha\beta) = \tau(\tau(\theta, \beta), \alpha)$ and invoking Lemma 4, we have

$$\frac{\partial}{\partial \alpha} [H(\tau(\theta, \alpha\beta))]_{\alpha=1} = -V(\tau(\theta, \beta)) \quad (41)$$

$$= -\beta^2 V(\tau(\theta, \beta) || \theta), \quad (42)$$

where (42) follows from Lemma 5 of [7]. Hence, by differentiating the above with respect to β at $\beta = 1$ and invoking Lemma 7, we arrive at the claim. ■

Proof of Theorem 5: The theorem is proved by invoking Lemmas 9 and 10, as follows:

$$H(\theta)S(\theta) < V^2(\theta) + \phi^2(1-\phi)(1-2\phi) \log^3 \left(\frac{1-\phi}{\phi} \right) \quad (43)$$

$$< V^2(\theta) + V(\theta)H(\theta) \quad (44)$$

$$< V^2(\theta) + 2V(\theta)H(\theta), \quad (45)$$

and hence θ satisfies the SEC. ■

Lemma 9. For any $\theta \in \Theta_2$, we have

$$H(\theta)S(\theta) < V^2(\theta) + \phi^2(1-\phi)(1-2\phi) \log^3 \left(\frac{1-\phi}{\phi} \right), \quad (46)$$

where $\phi := \min\{\theta_1, \theta_2\}$.

Proof: Let $\phi = \min\{\theta_1, \theta_2\}$. First note that by Lemma 11, we have

$$H(\theta) < \phi \log \frac{1}{\phi} + \phi. \quad (47)$$

Hence,

$$\begin{aligned} H(\theta)S(\theta) &< \phi^2(1-\phi)(1-2\phi)\log^3\left(\frac{1-\phi}{\phi}\right) \\ &+ \phi^2(1-\phi)(1-2\phi)\log^3\left(\frac{1-\phi}{\theta}\right)\log\left(\frac{1}{\phi}\right) \end{aligned} \quad (48)$$

$$\begin{aligned} &< \phi^2(1-\phi)(1-2\phi)\log^3\left(\frac{1-\phi}{\phi}\right) \\ &+ \phi^2(1-\phi)^2\log^4\left(\frac{1-\phi}{\phi}\right), \end{aligned} \quad (49)$$

where (49) follows from Lemma 12, completing the proof. ■

Lemma 10. For any $\theta \in \Theta_2$, we have

$$H(\theta)V(\theta) > \phi^2(1-\phi)(1-2\phi)\log^3\left(\frac{1-\phi}{\phi}\right), \quad (50)$$

where $\phi := \min\{\theta_1, \theta_2\}$.

Proof: For $\phi = \min\{\theta_1, \theta_2\}$, note that

$$H(\theta) > \phi \log \frac{1}{\phi}, \quad (51)$$

and hence

$$H(\theta)V(\theta) > \phi^2(1-\phi)\log^2\left(\frac{1-\phi}{\phi}\right)\log\left(\frac{1}{\phi}\right) \quad (52)$$

$$> \phi^2(1-\phi)(1-2\phi)\log^3\left(\frac{1-\phi}{\phi}\right), \quad (53)$$

where (53) follows from Lemma 12, completing the proof. ■

Lemma 11. For any $0 < x < 1$, we have

$$(1-x)\log\frac{1}{1-x} < x. \quad (54)$$

Proof: Note that as $x \rightarrow 0$ both sides are equal and the limit of their derivatives are equal as well, while the second derivative of the left hand side is equal to $-\frac{1}{1-x} < 0$ completing the proof. ■

Lemma 12. For any $0 < x < \frac{1}{2}$, we have

$$(1-2x)\log\frac{1}{x} < (1-x)\log\frac{1-x}{x}. \quad (55)$$

Proof: The proof is similar to that of Lemma 11. ■

Proof of Theorem 6: We proceed with the proof by construction. Let θ be such that

$$\theta_i = \begin{cases} (1 - \epsilon)/(|\mathcal{X}| - 1) & 1 \leq i \leq |\mathcal{X}| - 1 \\ \epsilon & i = |\mathcal{X}| \end{cases}. \quad (56)$$

Then, invoking Lemma 13, we can see that as $\epsilon \rightarrow 0$, for sufficiently small ϵ and $|\mathcal{X}| > 2$, we have

$$\frac{1}{2} \log(|\mathcal{X}| - 1) < H(\theta) < 2 \log(|\mathcal{X}| - 1), \quad (57)$$

$$\frac{1}{2} \epsilon \left(\log \frac{1}{\epsilon} \right)^2 < V(\theta) < \epsilon \left(\log \frac{1}{\epsilon} \right)^2, \quad (58)$$

$$\frac{1}{2} \epsilon \left(\log \frac{1}{\epsilon} \right)^3 < S(\theta) < \epsilon \left(\log \frac{1}{\epsilon} \right)^3. \quad (59)$$

Hence,

$$S(\theta)H(\theta) > \frac{1}{4} \epsilon \left(\log \frac{1}{\epsilon} \right)^3 \log(|\mathcal{X}| - 1) \quad (60)$$

$$> \epsilon^2 \left(\log \frac{1}{\epsilon} \right)^4 + 4\epsilon \left(\log \frac{1}{\epsilon} \right)^2 \log(|\mathcal{X}| - 1) \quad (61)$$

$$> V^2(\theta) + 2H(\theta)V(\theta). \quad (62)$$

where (61) holds for sufficiently small ϵ as long as $|\mathcal{X}| > 2$. Thus, θ does not satisfy the SEC, and the proof is complete. \blacksquare

Lemma 13. Let $\theta \in \Theta_{|\mathcal{X}|}$ be such that

$$\theta_i = \begin{cases} (1 - \epsilon)/(|\mathcal{X}| - 1) & 1 \leq i \leq |\mathcal{X}| - 1 \\ \epsilon & i = |\mathcal{X}| \end{cases}. \quad (63)$$

Then,

$$H(\theta) = (1 - \epsilon) \log(|\mathcal{X}| - 1) + h(\epsilon), \quad (64)$$

$$V(\theta) = \epsilon(1 - \epsilon) \left(\log \left(\frac{1 - \epsilon}{\epsilon} \right) - \log(|\mathcal{X}| - 1) \right)^2, \quad (65)$$

$$S(\theta) = \epsilon(1 - \epsilon)(1 - 2\epsilon) \left(\log \left(\frac{1 - \epsilon}{\epsilon} \right) - \log(|\mathcal{X}| - 1) \right)^3, \quad (66)$$

where $h(\epsilon)$ is the binary entropy function given by

$$h(\epsilon) := H(\epsilon, 1 - \epsilon) = \epsilon \log \frac{1}{\epsilon} + (1 - \epsilon) \log \frac{1}{1 - \epsilon}. \quad (67)$$

Proof: The calculation of $H(\theta)$ is straightforward by noting that this is a mixture of two uniform sources on alphabets of size $(|\mathcal{X}| - 1)$ and 1. To calculate $V(\theta)$, we have

$$\begin{aligned} V(\theta) &= (1 - \epsilon) \left(\log \frac{|\mathcal{X}| - 1}{1 - \epsilon} - (1 - \epsilon) \log(|\mathcal{X}| - 1) - h(\epsilon) \right)^2 \\ &\quad + \epsilon \left(\log \frac{1}{\epsilon} - (1 - \epsilon) \log(|\mathcal{X}| - 1) - h(\epsilon) \right)^2 \end{aligned} \quad (68)$$

$$\begin{aligned} &= (1 - \epsilon) \left(\epsilon \log(|\mathcal{X}| - 1) + \epsilon \log \frac{\epsilon}{1 - \epsilon} \right)^2 \\ &\quad + \epsilon \left(-(1 - \epsilon) \log(|\mathcal{X}| - 1) + (1 - \epsilon) \log \frac{1 - \epsilon}{\epsilon} \right)^2 \end{aligned} \quad (69)$$

$$= \epsilon(1 - \epsilon) \left(\log \frac{1 - \epsilon}{\epsilon} - \log(|\mathcal{X}| - 1) \right)^2. \quad (70)$$

Finally, to calculate $S(\theta)$, similarly to the calculations for $V(\theta)$, we get

$$\begin{aligned} S(\theta) &= (1 - \epsilon) \left(\epsilon \log(|\mathcal{X}| - 1) + \epsilon \log \frac{\epsilon}{1 - \epsilon} \right)^3 \\ &\quad + \epsilon \left(-(1 - \epsilon) \log(|\mathcal{X}| - 1) + (1 - \epsilon) \log \frac{1 - \epsilon}{\epsilon} \right)^3 \end{aligned} \quad (71)$$

$$= \epsilon(1 - \epsilon)(1 - 2\epsilon) \left(\log \frac{1 - \epsilon}{\epsilon} - \log(|\mathcal{X}| - 1) \right)^3, \quad (72)$$

establishing the claim. ■

Proof of Theorem 7: Let X be drawn from θ . Further, let

$$Y = \log \frac{1}{P(X)} - H(X).$$

Hence, by definition, $E[Y^3] = S(\theta)$ and $E[Y^2] = V(\theta)$. Then, the condition in (20) would ensure that $Y \in [-2, 2]$. Noting that the uniform distribution is excluded in $\Theta_{|\mathcal{X}|}$, and hence the varentropy is nonzero, we apply Lemma 14 (with $a = 2$) to obtain that

$$S(\theta) < 2V(\theta).$$

This is a sufficient condition for the SEC to hold, completing the proof. ■

Lemma 14. Let Y be a random variable supported on $[-a, a]$ for some $a > 0$. Further, let $E[Y] = 0$ and $E[Y^2] > 0$. Then,

$$\frac{E[Y^3]}{E[Y^2]} \leq a. \quad (73)$$

Proof: It is straightforward to show that $\frac{E[Y^3]}{E[Y^2]}$ is maximized if

$$p_y(y) = \begin{cases} \rho/2, & y = -a \\ 1 - \rho, & y = 0 \\ \rho/2, & y = a \end{cases},$$

for some $\rho > 0$, which in turn leads to $\frac{E[Y^3]}{E[Y^2]} = a$. ■

Proof of Corollary 8: First we show that the condition in (21) leads to the condition in (20), which follows from the following set of inequalities:

$$\begin{aligned} \max_{i \in [|\mathcal{X}|]} \left| \log \frac{1}{\theta_i} - H(\theta) \right| &\leq \max_{i \in [|\mathcal{X}|]} \left| \log \frac{1}{\theta_i} - \log |\mathcal{X}| \right| \\ &\quad + |\log |\mathcal{X}| - H(\theta)| \end{aligned} \tag{74}$$

$$\leq 2 \max_{i \in [|\mathcal{X}|]} \left| \log \frac{1}{\theta_i} - \log |\mathcal{X}| \right| \tag{75}$$

$$= 2, \tag{76}$$

where (74) follows Jensen's inequality and the convexity of the $|\cdot|$ operator, and (76) is a direct result of (21). Hence, the claim of Lemma 7 holds, which results in the claim of the theorem. ■

REFERENCES

- [1] J. L. Massey, "Guessing and entropy," in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*. IEEE, 1994, p. 204.
- [2] E. Arıkan, "An inequality on guessing and its application to sequential decoding," *Information Theory, IEEE Transactions on*, vol. 42, no. 1, pp. 99–105, 1996.
- [3] D. Malone and W. G. Sullivan, "Guesswork and entropy," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 525–526, Mar. 2004.
- [4] C. E. Pfister and W. G. Sullivan, "Renyi entropy, guesswork moments, and large deviations," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2794–2800, Nov. 2004.
- [5] M. M. Christiansen and K. R. Duffy, "Guesswork, large deviations, and shannon entropy," *Information Theory, IEEE Transactions on*, vol. 59, no. 2, pp. 796–802, 2013.
- [6] A. Beirami, R. Calderbank, M. Christiansen, K. Duffy, A. Makhdoumi, and M. Médard, "A geometric perspective on guesswork," in *53rd Annual Allerton Conference (Allerton)*, Oct. 2015.
- [7] A. Beirami, R. Calderbank, M. Christiansen, K. Duffy, and M. Médard, "A characterization of guesswork on swiftly tilting curves," *preprint*, 2017.
- [8] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.
- [9] A. Beirami, R. Calderbank, K. Duffy, and M. Médard, "Quantifying computational security subject to source constraints, guesswork and inscrutability," in *2015 IEEE International Symposium on Information Theory Proceedings (ISIT)*, Jun. 2015.
- [10] V. Strassen, "Asymptotische abschätzungen in shannons informations theorie," in *Trans. Third Prague Conf. Inf. Theory*, 1962, pp. 689–723.

- [11] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [12] E. Arıkan, “Varentropy decreases under the polar transform,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3390–3400, 2016.