To appear in *Topics in Cryptology - CT-RSA 2007, The Cryptographers' Track at the RSA Conference 2007*,**LNCS, Springer-Verlag.**

Selecting secure passwords

Eric R. Verheul

PricewaterhouseCoopers Advisory, Radboud University Nijmegen, Institute for Computing and Information Sciences, P.O. Box 85096, 3508 AB Utrecht, The Netherlands, `eric.verheul@[nl.pwc.com, cs.ru.nl]`

Abstract. We mathematically explore a model for the shortness and security for passwords that are stored in hashed form. The model is implicitly in the NIST publication [8] and is based on conditions of the Shannon, Guessing and Min Entropy. In addition we establish various new relations between these three notions of entropy, providing strong improvements on existing bounds such as the McEliece-Yu bound from [7] and the Min entropy lowerbound on Shannon entropy [3]. As an application we present an algorithm generating near optimally short passwords given certain security restrictions. Such passwords are specifically applicable in the context of one time passwords (e.g. initial passwords, activation codes).

1 Introduction

Consider the context of a computer system to which the logical access by a user (a human or another computer) is protected by a password. The threat that we consider is compromise of this password through one of the following two types of attack. In the *on-line guessing* type of attack, the attacker repeatedly makes guesses of the password, most likely first, and tests them by attempting to logon to the system. In our model the system has implemented "account lockout", locking the system after a certain number, say b, unsuccessful logon attempts which limits the effectiveness of the attack. In the *off-line guessing* type of attack, the attacker gets hold of some test-data from the system that enables him to test password guesses, most likely first, on his own systems. This information could for instance be a UNIX "passwd" file, a Windows SAM database or, more generally, the secure hash of a password. We distinguish two kinds of off-line attacks. In a *complete* attack the attacker is willing to take all the required computational effort to completely finish his attack algorithm thereby surely finding the password. In an *incomplete* attack the attacker is only willing to take a certain computational effort, a number of L guesses, in the attack, thereby finding the password only with a certain probability. To illustrate, suppose that an attacker has the SHA-1 hash of the password. If the attacker is willing to let the guess process run on a 1 GHz Pentium machine for a day this means that he is willing to perform about 2^{36} tries (cf. [2]); one might find it acceptable that the probability of success is at most 1%.

The central problem of this paper deals with choosing passwords that on the one hand have the functional requirement that they are "small" and on the other hand have the security requirement that they are "adequately" resistent against both on-line as off-line attacks, both complete as incomplete. Such passwords are specifically applicable in the context of one time passwords (e.g. initial passwords, activation codes).

Outline of the paper
In Section 2 we describe the mathematical model we use for the central problem in this paper. In this model the Shannon entropy is taken as a measure for "smallness" of passwords, the Guessing entropy [6] as a measure for security of passwords against complete off-line attacks and the Min entropy (cf. [3]) as a measure for security of passwords against incomplete off-line attacks. In Section 3 we discuss and apply some techniques for calculating extreme points of convex sets. In Section 4 we present general, new relations between the three types of entropy. This provides both strong improvements on the McEliece-Yu bound from [7] and the Min entropy lowerbound [3] on Shannon entropy. In Section 5 we arrive at a new lower bound on the Shannon entropy of distributions, i.e. on the minimal length of the corresponding passwords, given restrictions on the Guessing and Min entropy. As an application we present in Section 6 an algorithm generating near optimally short passwords under these conditions that are specifically applicable in the context of one time passwords (e.g. initial passwords, activation codes). Finally Section 7 contains the conclusion of this paper and open problems.

Related work
NIST Special Publication 800-63 [8] provides technical guidance in the implementation of electronic authentication, including passwords. The implicitly used mathematical model for security of passwords is similar to ours but the publication does not fully mathematically explore this model. Massey [6], cf. [5], shows that the entropy H for a password distribution is upper bounded in terms of its Guessing entropy α by $H \leq \log_2(e\cdot\alpha-1)$. Note that Massey's bound is independent of the number of passwords n. By a counterexample it is indicated in [6] that no interesting lower bound on the entropy of a password exists in terms of the Guessing entropy alone. McEliece and Yu [7] show that $H \geq \frac{2\log_2(n)}{n-1}(\alpha-1)$ indicating that such lower bounds exist if one takes into account the number of passwords n. It is well-known that the Shannon entropy is lower bounded by the Min entropy (cf. [3]), i.e., independent of n. Massey's bound can also be formulated as a lower bound on the Guessing entropy in terms of the Shannon entropy; in [1] Arikan provides another lower bound on the Guessing entropy in terms of the l^p-norm of the underlying probability distribution.

2 The mathematical model for secure passwords

In this section we describe our mathematical model for secure passwords selections. Further motivation of the model is placed in Appendix A. We assume that passwords correspond to a finite variable X with a discrete distribution $(p_1,\ldots,p_n)$ on n points (the number of possible passwords). That is, each $p_i \geq 0$

and they sum up to 1. To facilitate easy notation and formulae we assume throughout this paper that the probabilities p_i are denoted in a decreasing form, i.e., $p_1 \geq p_2 \ldots \geq p_n \geq 0$. The size of passwords is measured in our model by the *(Shannon) Entropy* $H(X)$ (or simply H) which is given by

$$H(X) = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i),$$

and where we use the usual convention that $0 \cdot \log_2(0) = 0$. Our choice is motivated by the fact that the entropy corresponds with the average size of passwords in bits using an optimal coding for these passwords, most notably a coding based on a Huffman encoding tree [4]. The resistance against complete off-line attacks we measure by the *Guessing entropy*, cf. [6], denoted by $G(X)$ or simply α, given by

$$G(X) = \sum_{i=1}^{n} i \cdot p_i.$$

This relates to the expected number of tries for finding the password using an optimal strategy, i.e. trying the most likely keys first. Perhaps surprising, a large Guessing entropy by itself is not a sufficient guarantee for a secure passwords distribution. In [6] an example is given of a family of distributions (see also Section 4) that all have a fixed Guessing entropy but that have a highest probability that increases to one (and the entropy decreases to zero). That is, the probability that the first guess is successful goes to one implying that these distributions are certainly not "secure". From this example it is indicated that a viable security model also needs to take into account resistance against incomplete attacks. In our model, we measure this by the so-called *Min Entropy*, $H_\infty(X)$ or simply H_∞ given by $-\log_2(p_1)$, cf. [3]. If the Min entropy is sufficiently large, or equivalently p_1 sufficiently small, then one can assure that $\sum_{i=1}^{L} p_i \leq L \cdot p_1$ is sufficiently small too.

The resistance against on-line attacks is directly related to the probability of success in the optimal strategy, i.e. trying the b most likely keys. As typically the acceptable number b will be much smaller than the acceptable number L of an incomplete attack, we will only impose conditions on the effectiveness on the latter kind of attack. The central problem of this paper can now be mathematically formulated as follows: given lower bounds on the Guessing and Min entropy (or equivalently an upper bound on p_1) what is the minimal Shannon entropy possible and how can one calculate and apply minimal distributions?

3 Preliminaries

3.1 Extreme points of ordered distributions with fixed Guessing entropy

We recall (cf. [9, p. 241]) that a point x of a convex set C is *extreme* if it is not an interior point of any line segment lying in K. Thus x is extreme iff whenever $x = \lambda y + (1-\lambda)z$ with $0 < \lambda < 1$ we have $y \notin C$ or $z \notin C$. It is a direct consequence of the Krein-Milman theorem (cf. [9, p. 242]) that any closed, bounded convex set in $\mathbb{R}^n$ for some natural n is the convex hull of its extreme points.

Let r, s, n be natural numbers and let $f_1, ..., f_r$ and $F_1, ..., F_s$ be (linear) functionals on $\mathbb{R}^n$ and let $\delta_1, ..., \delta_r, \theta_1, ..., \theta_s \in \mathbb{R}$. The set $C \subset \mathbb{R}^n$ is defined by

$$C = \{x \in \mathbb{R}^n | f_i(x) = \delta_i \text{ for } i = 1, 2, ..., r \text{ and } F_j(x) \geq \theta_j \text{ for } j = 1, 2..., s\}.$$

Clearly, C is a closed convex set but is not necessarily bounded. Equations of type $f_i(x) = \delta_i$ we call *defining hyperplanes* and equations of type $F_j(x) \geq \theta_j$ we call *defining halfspaces*. We call a point x in C a *minimal intersection point* if

$$\{x\} = \cap_{i=1}^{r} f^{-1}(\delta_i) \bigcap \cap_{j \in S} F^{-1}(\theta_j) \tag{1}$$

for some subset S of $\{1, ..., s\}$. Determining the elements in (1) amounts to solving n variables based on $r + \|S\|$ equations as indicated in (1). Minimal intersection points then coincide with unique solutions x to such sets of equations that also satisfy the other conditions, i.e. lie in the remaining defining halfspaces. If each subset of n functionals in $\{f_1(\cdot), ..., f_r(\cdot), F_1(\cdot), ..., F_s(\cdot)\}$ is linearly independent (and so $r \leq n$ in particular), then one only needs to look at subset S of size $n - r$. The following result, using the notation of above, can be considered as part of the mathematical "folklore", cf. [11]. We provide proofs in Appendix B.

Theorem 1 *If C is bounded then the extreme points of C are precisely the minimal intersection points and C is the convex hull of the minimal intersection points.*

Let n be a natural number and α be a real number and let

$$C_{n,\alpha} = \{(p_1, ..., p_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} ip_i = \alpha, p_1 \geq p_2 ... \geq p_n \geq 0\}.$$

One can easily verify that $C_{n,\alpha} \neq \emptyset$ iff $1 \leq \alpha \leq (n+1)/2$ so from now on we implicitly assume that α satisfies the condition $1 \leq \alpha \leq (n+1)/2$. It is easily verified that $C_{n,1} = \{(1, 0, \ldots, 0)\}$ and $C_{n,(n+1)/2} = \{(1/n, 1/n, \ldots, 1/n)\}$.

Theorem 2 *The set $C_{n,\alpha}$ is a closed, bounded the convex set and is the hull of its extreme points. These extreme points take the form $X_{j,k,n}$ for integers j, k satisfying $1 \leq j \leq 2\alpha - 1 \leq k \leq n$ and*

$$\begin{array}{rcccccccccc} X_{j,k,n} = (& a_{j,k,n}, & a_{j,k,n}, & \cdots & a_{j,k,n}, & b_{j,k,n}, & \cdots & b_{j,k,n}, & 0, & \cdots & 0) \\ & & & & \uparrow & \uparrow & & \uparrow & \uparrow & & \\ & 1, & 2, & \cdots & j, & j+1, & \cdots & k, & k+1, & \cdots & n, \end{array}$$

where

$$a_{j,k,n} = \frac{-2\alpha + 1 + j + k}{j \cdot k}; b_{j,k,n} = \frac{2\alpha - (j+1)}{k(k-j)},$$

and where we define $b_{j,k,n} = 1/(2\alpha - 1)$ for $j = 2\alpha - 1 = k$ (which can only occur when $2\alpha - 1$ is an integer).

Proof: See Appendix B. $\square$

Note that if in the previous theorem $2\alpha-1$ is an integer then all points of type $X_{j,2\alpha-1,n}$ for $1 \leq j \leq 2\alpha-1$ are equal to the point whose first $2\alpha-1$ coordinates are equal to $1/(2\alpha-1)$ and the remaining ones are zero. We note that from the previous theorem it simply follows that the set $C_{n,\alpha}$ has more than one point if $\alpha < (n+1)/2$ and $n \geq 3$. So, for $n \geq 3$ it follows $C_{n,\alpha} = \{(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})\}$ iff $|C_{n,\alpha}| = 1$.

In practice often a certain number, say d, of the highest probabilities coincide. For instance when a dictionary is used and the d most likely passwords are chosen uniformly from a dictionary of size $d \leq n$. The set of such distributions takes the following form:

$$C_{n,\alpha,d} = \{(p_1, ..., p_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} i \cdot p_i = \alpha, \\ p_1 = p_2 \ldots = p_d \geq p_{d+1} \ldots \geq p_n \geq 0\}$$

We usually write $C_{n,\alpha}$ for $C_{n,\alpha,1}$. The following two results state some of the immediate properties of $C_{n,\alpha}$ and $C_{n,\alpha,d}$.

Proposition 1 *$C_{n,\alpha,d} \neq \emptyset$ iff $C_{n,\alpha} \neq \emptyset$ and $d \leq 2\alpha - 1$ iff $1 \leq \alpha \leq (n+1)/2$ and $d \leq 2\alpha - 1$.*

Proof: See Appendix B. $\square$

Proposition 2 *We use the terminology and conditions of Theorem 2. The extreme points of $C_{n,\alpha}$ are the points $X_{j,k,n}$ satisfying $d \leq j \leq 2\alpha - 1 \leq k \leq n$.*

Proof: See Appendix B. $\square$

3.2 Useful formulae

We use the terminology of Theorem 2. It is convenient to introduce the function $G(x, y, z)$ with domain $\{(x, y, z) \in \mathbb{R}^3 \mid 0 < x \leq y \leq z\}$ given by

$$G(x, y, z) = -\left(\frac{(-y + x + z)}{z} \log_2(\frac{-y + x + z}{x \cdot z}) \right. \\ \left. + \frac{(y - x)}{z} \log_2(\frac{y - x}{z(z - x)})\right),$$

if $x < y \leq z$ and $G(y, y, z) = \log_2(y)$ and $G(y, y, y) = \log_2(y)$. Note that $G(x, z, z) = \log_2(z)$, $\lim_{x \uparrow y} G(x, y, z) = G(y, y, z)$ and that $G(j, 2\alpha - 1, k) = H(X_{j,k,n})$. Also note that values of $G(\cdot)$ are easily calculated. As usual we denote the entropy function on two points by $h(\cdot)$, i.e., $h(p) = -p \log_2(p) - (1-p) \log_2(1-p)$.

A real valued function $G(\cdot)$ from a convex set $C \subset \mathbb{R}^n$ is called *convex* (respectively *concave*) if $G(\lambda x + (1-\lambda)y) \leq \lambda G(x) + (1-\lambda)G(y)$ (respectively $G(\lambda x + (1-\lambda)y) \geq \lambda G(x) + (1-\lambda)G(y)$ for all $x, y \in C$ and $0 \leq \lambda \leq 1$. Note that $G(\cdot)$ is convex iff $-G(\cdot)$ is concave. If $G(\cdot)$ is a twice-differentiable function in a single real variable (i.e., $C \subset \mathbb{R}$) then $G(\cdot)$ is convex (respectively concave) if $G^{''} \geq 0$ (respectively $G^{''} \leq 0$) on C. The proofs of the following two lemmas are straightforward verifications.

Lemma 1 *For $1 \leq x \leq y \leq z$ the following holds.*

1. $G(x, y, z) - \log_2(x) = G(1, y/x, z/x)$
2. $G(1, y, z) = h(\frac{y-1}{z}) + \frac{y-1}{z}\log_2(z-1)$

Lemma 2

1. *For fixed x, z, the function $[x, z] \ni y \to G(x, y, z)$ is concave.*
2. *For fixed y, z, the function $[1, y] \ni x \to G(x, y, z)$ is concave.*
3. *For fixed x, y, the function $[y, \infty) \ni z \to G(x, y, z)$ increases to its maximum and is then decreasing.*

Proposition 3 $G(x, y, z) \geq \log_2(x) + \frac{y-x}{z-x}\log_2(z/x)$

Proof: By Lemma 2, for fixed x, z the function $[x, z] \ni y \to G(x, y, z)$ is concave. Note that for $y = x$ this function takes the value $\log_2(x)$ and for $y = x$ it takes the value $\log_2(z)$. If we write $y = (1-\lambda)x + \lambda z$ it follows that $\lambda = (y-x)/(z-x)$. From concavity it now follows

$$G(x, y, z) \geq (1-\lambda)\log_2(x) + \lambda \log_2(z) = \log_2(x) + \lambda \log_2(z/x),$$

from which the proposition follows. □

4 Relations between entropies

As explained in the introduction, we let passwords correspond to a finite variable X with a discrete distribution $(p_1, \ldots, p_n)$ on n points that we assume to be ordered, i.e., $p_1 \geq p_2 \ldots \geq p_n \geq 0$. The following inequalities hold, providing lower and upper bounds for the highest probability, i.e. p_1, of points in $C_{n,\alpha,d}$ in terms of α and n.

Theorem 3 *For $(p_1, \ldots, p_n) \in C_{n,\alpha,d}$ the following inequalities hold*

$$\frac{1}{2\alpha-1} \leq \frac{-2\alpha+1+\lfloor 2\alpha-1 \rfloor}{(\lfloor 2\alpha-1 \rfloor)(\lceil 2\alpha-1 \rceil)} + \frac{1}{\lfloor 2\alpha-1 \rfloor} \leq p_1 \leq 1/d + 1/n - \frac{2\alpha-1}{n}.$$

Proof: If $d = 1$ then from Theorem 2 it follows that the extreme points of $C_{n,\alpha,d}$ are of type $X_{j,k,n}$ with $1 \leq j \leq 2\cdot\alpha - 1 \leq k \leq n$. It more generally follows from Proposition 2 that for $d > 1$, the extreme points of $C_{n,\alpha,d}$ are of type $X_{j,k,n}$ with

$1 \leq j \leq 2 \cdot \alpha - 1 \leq k \leq n$.[1] It follows that $(p_1, \ldots, p_n)$ is in the convex hull of the extreme points $X_{j,k,n}$ with $d \leq j \leq 2 \cdot \alpha - 1 \leq k \leq n$.

Note that for fixed k the formula of the first coordinate of $X_{j,k,n}$, i.e., $a_{j,k,n}$, is decreasing in j. As the smallest permissible j equals d and the largest permissible equals $\lfloor 2\alpha - 1 \rfloor$, it follows that

$$\min\{a_{\lfloor 2\alpha-1 \rfloor,k,n} \mid 2\alpha - 1 \leq k \leq n\} \leq p_1 \leq \max\{a_{d,k,n} \mid 2\alpha - 1 \leq k \leq n\} \quad (2)$$

Also note that for fixed j the formula of the first coordinate of $X_{j,k,n}$, i.e., $a_{j,k,n}$, is increasing in k. This means that the left hand side of (2) is equal to $a_{\lfloor 2\alpha-1 \rfloor, \lceil 2\alpha-1 \rceil, n}$, which is easily seen equal to

$$\frac{-2\alpha + 1 + \lfloor 2\alpha - 1 \rfloor}{(\lfloor 2\alpha - 1 \rfloor)(\lceil 2\alpha - 1 \rceil)} + \frac{1}{\lfloor 2\alpha - 1 \rfloor}.$$

That this expression is greater or equal to $1/(2\alpha - 1)$ follows from the easily verified inequality $x - \lfloor x \rfloor / (\lfloor x \rfloor \cdot \lceil x \rceil) + 1/\lfloor x \rfloor \geq 1/x$ for any $x \geq 0$. This concludes the proof for the second equality of the result. Similarly, the right hand side of (2) is equal to $a_{d,n,n}$ which is easily seen equal to $1/d + 1/n - \frac{2\alpha-1}{n}$ completing the proof of the theorem. $\square$

As the function $-\log_2(\cdot)$ is decreasing, the bounds in the previous result can easily transformed in lower- and upperbounds for the Min entropy H_∞ in terms of α and n. The following result is a direct consequence; the right hand inequality also follows from a standard concavity result.

Corollary 1 $-\log_2(1 - \frac{2(\alpha-1)}{n}) \leq H_\infty \leq \log_2(2\alpha - 1)$

The next result enables the precise calculation of the minimum entropy on $C_{n,\alpha,d}$ that we denote by $M_{n,\alpha,d}$.

Theorem 4 *The following hold:*

1. $M_{n,\alpha,d} = \min\{G(j, 2\alpha - 1, k) \mid j \in \{d, \lfloor 2\alpha - 1 \rfloor\}, k \in \{\lceil 2\alpha - 1 \rceil, n\}\}$.
2. $M_{n,\alpha,d} \geq \min\{\log_2(2\alpha-1), G(d, 2\alpha-1, n)\} = \min\{\log_2(2\alpha-1), h(\frac{2\alpha-1-d}{n}) + \frac{2\alpha-d}{n}\log_2(n/d - 1)\}$ *with equality if* $2\alpha - 1$ *is an integer.*

Proof: We prove the two parts of the theorem simultaneously. As the entropy function is concave, the minimum of the entropy function on $C_{n,\alpha,d}$ is the minimum the entropy achieves on its extreme points, i.e., the points of type $X_{j,k,n}$ with $d \leq 2\alpha - 1 \leq k \leq n$ (cf. Proposition 2), that is:

$$M_{n,\alpha,d} = \min\{G(j, 2\alpha - 1, k) \mid k, j \in \mathbb{N}, d \leq j \leq 2\alpha - 1 \leq k \leq n\}$$

From the concavity of the function $j \to G(j, 2\alpha - 1, k)$, i.e. Lemma 2, follows

$$M_{n,\alpha,d} = \min\{G(j, 2\alpha - 1, k) \mid j \in \{d, \lfloor 2\alpha - 1 \rfloor\}, k \in \mathbb{N}, 2\alpha - 1 \leq k \leq n\} \quad (3)$$

[1] We distinguish between $d = 1$ and $d > 1$ to avoid a circular reasoning in the proofs of Propositions 1, 2 and Theorem 3.

$$M_{n,\alpha} \geq \min\{G(j, 2\alpha-1, k) \mid j \in \{d, 2\alpha-1\}, k \in \mathbb{N}, 2\alpha-1 \leq k \leq n\} \qquad (4)$$

with equality if $2\alpha-1$ is an integer. Finally, from equality (3) and the third part of of Lemma 2 we arrive at the first part of the theorem.

As $G(2\alpha-1, 2\alpha-1, n) = G(d, 2\alpha-1, 2\alpha-1) = G(2\alpha-1, 2\alpha-1, 2\alpha-1) = \log_2(2\alpha-1)$ inequality (4) implies that

$$\begin{aligned} M_{n,\alpha,d} &\geq \min\{G(j, 2\alpha-1, k) \mid j \in \{d, 2\alpha-1\}, k \in \{2\alpha-1, n\}\} \\ &= \min\{\log_2(2\alpha-1), G(d, 2\alpha-1, n)\} \end{aligned}$$

with equality if $2\alpha-1$ is an integer. The second part of the theorem now follows from combining the two formulae from Lemma 1. □

From Theorem 4 it follows that $M_{n,\alpha,d}$ is asymptotically equal to $G(d, 2\alpha, n)$, i.e. to the entropy of the distribution the $X_{d,n,n}$ that goes to $\log_2(d)$. For $d = 1$ this sequence actually forms the counterexample in [6] that no (interesting) lower bound on the entropy exists in terms of the Guessing entropy alone.

Theorem 4 also enables to determine that perhaps contrary to popular belief, the formula $\log_2(2\alpha-1) \leq H$ is actually only true for $n \leq 6$. Indeed, it is easily verified that the graph of the function $h_n : [1, (n+1)/2] \ni \alpha \to G(1, 2\alpha-1, n)$ lies under the graph of $[1, (n+1)/2] \ni \alpha \to \log_2(2\alpha-1)$ for $n = 1, 2, ..., 6$. From the second part of Theorem 4 it now follows that the formula is true for $n \leq 6$. That this formula does not hold for $n \geq 7$ also follows from the second part of Theorem 4 as $h_7(1.5) = 0.960953 < 1 = \log_2(2\alpha-1)$.

Theorem 5

$$M_{n,\alpha,d} \geq \log_2(d) + \frac{2\alpha-1-d}{n-d} \log_2(n/d)$$

This lowerbound on $M_{n,\alpha,d}$ is weaker than the lowerbound from the second part of Theorem 4. Moreover, both sides of the inequality are asymptotically equivalent in n.

Proof: Consider the following inequalities.

$$\begin{aligned} M_{n,\alpha,d} &\geq \min\left(G(d, 2\alpha-1, n), \log_2(2\alpha-1)\right) \\ &\geq \min\left(\log_2(d) + \frac{2\alpha-1-d}{n-d}\log_2(n/d), \log_2(2\alpha-1)\right) \\ &\geq \log_2(d) + \frac{2\alpha-1-d}{n-d}\log_2(n/d), \end{aligned}$$

The first inequality is the second part of Theorem 4 and the second inequality follows from Proposition 3. For the last inequality; note that the function $[2\alpha-1, \infty) \ni n \to \log_2(d)+\frac{2\alpha-1-d}{n-d}\log_2(n/d)$, is decreasing. As this function converges to the value $\log_2(2\alpha-1)$ for $n \downarrow 2\alpha-1$, the last inequality follows and thereby the first two parts of the theorem.

With respect to the last part of the theorem; the sequence of distributions $X_{d,n,n}$ all have Guessing entropy equal to α and it is easily seen that their Shannon entropies converge to $\log_2(d)$. $\square$

The previous result with $d = 1$ is an extension of the McEliece-Yu bound from [7]. Its proof also shows that the lowerbound in the second part of Theorem 4 for $d = 1$ provides a stronger bound on $M_{n,\alpha,1}$ than the McEliece-Yu bound. It is easily shown that the difference between these bounds is about $h(2(\alpha-1)/n)$, about one in practice.

5 Secure password distributions

Let $(p_1, ..., p_n) \in C_{n,\alpha}$ and let $0 < \delta \leq 1$. Then $C_{n,\alpha,\delta}$ is the set:

$$\{(p_1, ..., p_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} i \cdot p_i = \alpha, \delta \geq p_1 \geq p_2 ... \geq p_n \geq 0\}.$$

From the proof of Theorem 3 it follows that the extreme point $X_{\lfloor 2\alpha-1\rfloor,\lceil 2\alpha-1\rceil,n} \in C_{n,\alpha}$ has a minimal first coordinate, namely $a_{\lfloor 2\alpha-1\rfloor,\lceil 2\alpha-1\rceil,n}$. So $C_{n,\alpha,\delta} \neq \emptyset$ iff $C_{n,\alpha} \neq \emptyset$ and $a_{\lfloor 2\alpha-1\rfloor,\lceil 2\alpha-1\rceil,n} \leq \delta$. Or in other words that

$$\frac{-2\alpha + 1 + \lfloor 2\alpha - 1\rfloor}{(\lfloor 2\alpha - 1\rfloor)(\lceil 2\alpha - 1\rceil)} + \frac{1}{\lfloor 2\alpha - 1\rfloor} \leq \delta \text{ and } \alpha \leq (n+1)/2. \tag{5}$$

Clearly, the fact that $C_{n,\alpha,\delta}$ is non-empty does not imply that it contains an element with first coordinate equal to δ. We call δ *admissible* if there is such an element. It simply follows from the proof of Theorem 3 that δ is admissible iff

$$\frac{-2\alpha + 1 + \lfloor 2\alpha - 1\rfloor}{(\lfloor 2\alpha - 1\rfloor)(\lceil 2\alpha - 1\rceil)} + \frac{1}{\lfloor 2\alpha - 1\rfloor} \leq \delta \leq 1 - \frac{2(\alpha-1)}{n}. \tag{6}$$

The following theorem discusses the extreme points of $C_{n,\alpha,\delta}$ and how to calculate them. For $v \in \mathbb{R}^m$ we let $(v)_1$ denote the first coordinate of v.

Theorem 6 *The set of points E'*

$$\{\lambda X_{j_1,k_1,n} + (1-\lambda)X_{j_2,k_2,n} \mid j_1 = j_2 \textit{ or } k_1 = k_2, \lambda \in [0,1] : \\ \lambda(X_{j_1,k_1,n})_1 + (1-\lambda)(X_{j_2,k_2,n})_1 = \delta\} \cup \{f \in E \mid (f)_1 \leq \delta\}$$

is finite and its convex hull spans $C_{n,\alpha,\delta}$. In particular, all extreme points of $C_{n,\alpha,\delta}$ are in E'.

Proof: See Appendix B. $\square$

Let $H(n, \alpha, \delta)$ denote $\min\{H(c) \mid c \in C_{n,\alpha,\delta}\}$. The following immediate result shows how $H(n, \alpha, \delta)$ can be precisely calculated.

Theorem 7 *$H(n, \alpha, \delta)$ is equal to the minimum value that the entropy takes on the set E' defined in Theorem 6, which is the minimum of at most $\lfloor 2\alpha - 1\rfloor * \lfloor n - 2\alpha - 1\rfloor$ real numbers.*

We extend the meaning of

$$a_{j,k,n} = \frac{-2\alpha + 1 + j + k}{j \cdot k}$$

from Theorem 2 to include any real $0 < j \leq 2\alpha - 1 \leq k \leq n$. The function $(0, 2\alpha - 1] \ni j \to a_{j,k,n}$ is decreasing and takes as an image the segment $[1/(2\alpha - 1), \infty)$. The inverse of this function is a function $g_k^\alpha : [1/(2\alpha - 1), \infty) \to (0, 2\alpha - 1]$ given by $g_k^\alpha(x) = \frac{k - 2\alpha + 1}{kx - 1}$.

The following is the main result of this section. The idea of calculating an lower bound on the Shannon entropy given a Guessing entropy and α upper-bound δ on the highest probability occurring is simple: just find the real number j such that the "virtual" extreme point $X_{j,n,n}$ has a first probability $a_{j,n,n}$ equal to δ. The lowerbound on the Shannon entropy is then the minimum of the entropy of the "virtual" extreme point and $\log_2(2\alpha - 1)$. The proof of Theorem 8 is placed in Appendix C.

Theorem 8 *Let α be fixed, $1 \leq 2\alpha - 1 \leq n$, and let δ be such that $C_{n,\alpha,\delta} \neq \emptyset$ (so in particular $\delta \geq 1/(2\alpha - 1)$, then*

$$H(n, \alpha, \delta) \geq \min(G(g_n^\alpha(\delta), 2\alpha - 1, n), \log_2(2\alpha - 1)).$$

Moreover, the sequence $\{\min(G(g_n^\alpha(\delta), 2\alpha - 1, n), \log_2(2\alpha - 1))\}_n$ is decreasing and converges to $-\log_2(\delta)$.

The following result is a consequence of Theorem 8 in an analogous fashion as Theorem 5 is a consequence of Theorem 4.

Theorem 9 *Let α be fixed, $1 \leq 2\alpha - 1 \leq n$, and let δ be such that $C_{n,\alpha,\delta} \neq \emptyset$ (so in particular $\delta \geq 1/(2\alpha - 1)$, then*

$$H(n, \alpha, \delta) \geq \log_2(g_n^\alpha(\delta)) + \frac{2\alpha - 1 - g_n^\alpha(\delta)}{n - g_n^\alpha(\delta)} \log_2(n/g_n^\alpha(\delta)). \tag{7}$$

This lowerbound on $H_{n,\alpha,\delta}$ is weaker than the lowerbound from the second part of Theorem 8. Moreover, the sequence $\{\log_2(g_n^\alpha(\delta)) + \frac{2\alpha - 1 - g_n^\alpha(\delta)}{n - g_n^\alpha(\delta)} \log_2(n/g_n^\alpha(\delta))\}_n$ is decreasing and converges to $-\log_2(\delta)$.

Proof Define the following function

$$F : (a, \infty) \ni n \to \log_2(g_n^\alpha(\delta)) + \frac{2\alpha - 1 - g_n^\alpha(\delta)}{n - g_n^\alpha(\delta)} \log_2(n/g_n^\alpha(\delta)).$$

The theorem follows from the following properties: a) $\lim_{n \downarrow 2\alpha - 1} F(n) = \log_2(2\alpha - 1)$, b) $\lim_{n \to \infty} F(n) = -\log_2(\delta)$ and c) $F(\cdot)$ is decreasing. Indeed, the proof of the first parts of the theorem follow from properties a) and c) similar to the proof of Theorem 5. The last part of the theorem follows from property b).

The first two properties are straightforward verifications. For a proof of the last property; if we denote $2\alpha - 1$ by a then the derivative of F to n equals:

$$\frac{(-1+\delta a)}{\log(2)} \cdot \left(n^2\delta - 2\,n + a\right)^{-2} \left((n^2\delta - a) \cdot \log(\frac{n-a}{(n\delta - 1)n}) + 2(n^2x - 2n + a) \right)$$

It suffices to show that the far right factor in this expression is negative. A straightforward verification shows that the far right factor divided by the expression $(n\delta - 1)n$ is equal to

$$(1+B)\log(B) + 2(1-B) \tag{8}$$

where $B = (n-a)/(n\delta - 1)n \in (0,1]$. It simply follows that expression (8) is ≤ 0, showing the last property. $\square$

Corollary 2 *Let $(p_1, p_2, ..., p_n)$ be an (ordered) password distribution with Shannon entropy H, Guessing entropy α and Min entropy H_∞ then*

$$\begin{aligned} H &\geq \min(G(g_n^\alpha(p_1), 2\alpha - 1, n), \log_2(2\alpha - 1)) \\ &\geq \log_2(g_n^\alpha(p_1)) + \frac{2\alpha - 1 - g_n^\alpha(p_1)}{n - g_n^\alpha(p_1)} \log_2(n/g_n^\alpha(p_1)) \geq H_\infty. \end{aligned}$$

We make some remarks on Theorem 8. If we fill in the largest possible admissible δ in Theorem 8, i.e., $\delta = 1 - 2(\alpha - 1)/n$, we obtain the second part of Theorem 4 for $d = 1$ which itself an improvement of the McEliece-Yu bound by Theorem 5. The bound in Theorem 8 is strong in the sense that for δ that equal the first coordinate of an extreme point of type $X_{j,n,n}$, i.e. $\delta = a_{j,n,n}$ equality in Theorem 8 holds provided $H(X_{j,n,n}) \leq \log_2(2\alpha - 1)$. In Appendix D it is further indicated that the bound in Theorem 8 is strong and that taking the minimum with $\log_2(2\alpha - 1)$ cannot be relaxed. It is also indicated that the distributions $X_{j,n,n}$ are in fact "local" minima.

6 Selecting near optimal secure passwords

The strongness discussed above of Theorem 8 gives rise to the following algorithm, providing a near optimal password distribution with minimal Shannon entropy in our security model. In this algorithm we assume that the Guessing entropy α is an integer and that the bound on the highest occurring probability δ is of the form $1/D$ for some natural number D. These are very mild restrictions. For the existence of such distributions (cf. Theorem 3) one requires that $1/(2\alpha - 1) \leq \delta$. If the latter equality holds, the only distribution satisfying is the uniform one on $2\alpha - 1$ points and the (minimal) Shannon entropy equals the Guessing entropy. So we assume that $1/(2\alpha - 1) < \delta$, i.e. $D < 2\alpha - 1$. As $\lim_{n\to\infty} g_n(1/D) \uparrow D$ it also follows that

$$\lim_{n\to\infty} G(g_n(1/D), 2\alpha - 1, n) = \lim_{n\to\infty} G(D, 2\alpha - 1, n) = \log_2(D).$$

It now follows from Theorem 8 that when n grows to infinity, the minimum Shannon entropy $H(n,\alpha,\delta)$ decreases to $\log_2(D)$. For two reasons the distribution $X_{D,n,n}$ is an obvious choice for a finite approximation of this minimum. Firstly, its Shannon entropy converges to this minimum. Secondly, the highest probability occurring in $X_{D,n,n}$, i.e., $a_{D,k,n}$, is less than δ but can be taken arbitrarily close it. Moreover, as discussed at the end of Section 5, for n large enough the distribution $X_{D,n,n}$ establishes the minimum Shannon entropy in its own "class", i.e., the distributions on n points with Guessing entropy equal to α and highest probability $\leq a_{D,n,n}$.

The distributions of type $X_{D,n,n}$ also have a simple form: the first D coordinates are equal to $a_{D,n,n}$ and the remaining $n-D$ coordinates are equal to $b_{D,n,n}$. This makes generation of passwords in accordance with this distribution quite convenient by using a Huffman tree as follows. First generate a '0' with probability $P_{min} = D * a_{D,n,n}$ and a '1' with probability $P_{max} = (n-D) * b_{D,n,n}$. If 0 is generated, generate a random string of size D bits and concatenate it with '0', if '1' is generated then generate a random string of size $(n-D)$ bits and concatenate it with '1'. One can easily verify that the average size of such generated strings is at most the Shannon entropy of $X_{D,n,n}$ plus one bit. By increasing n one obtains password generation methods with average bit length arbitrarily close to $\log_2(D)$ whereby with a small probability (i.e., $(n-D) * b_{D,n,n}$ decreasing to zero) large passwords will occur of size $\log_2(n-D) \approx \log_2(n)$. In the table below we have placed some of the characteristic figures for $\alpha = 2^{64}$ and $\delta = 2^{-40}$.

$\log_2(n)$	$-\log_2(a_{D,n,n})$	Average pwd length	Min length	Max length	P_{min}	P_{max}
65.0	65.00	65.00	40.0	65.0	2.98E-08	1.00E+00
65.5	41.77	58.90	40.0	65.5	2.92E-01	7.07E-01
66.0	41.00	54.00	40.0	66.0	5.00E-01	5.00E-01
66.5	40.62	50.30	40.0	66.5	6.46E-01	3.53E-01
67.0	40.41	47.56	40.0	67.0	7.50E-01	2.50E-01
67.5	40.28	45.53	40.0	67.5	8.23E-01	1.76E-01
68.0	40.19	44.04	40.0	68.0	8.75E-01	1.25E-01
68.5	40.13	42.95	40.0	68.5	9.11E-01	8.83E-02
69.0	40.09	42.14	40.0	69.0	9.37E-01	6.25E-02
69.5	40.06	41.56	40.0	69.5	9.55E-01	4.41E-02
70.0	40.04	41.13	40.0	70.0	9.68E-01	3.12E-02

If one applies this password generation method in a context where the system generates user passwords to be used repeatedly by the user, the user will be inclined to have changed the issued large password until the system proposes a small password. This of course undermines the security assumptions of the system. Also when using passwords repeatedly, it is important that they are easily memorable which the generated passwords in their current form are not. Consequently the password generation method described is only practically applicable when the passwords generated are One Time Passwords (OTPs). OTPs arise in many applications such as in activation codes for digital services (e.g. prepaid mobile credit typically stored on a scratch card). Also initial computer passwords supplied by the IT department of an organization can be considered to be OTPs.

7 Conclusion

We have presented a mathematical model for secure passwords and we have presented an algorithm providing near-optimal distributions in this model as well as a simple algorithm generating binary passwords accordingly. Such algorithms are specifically applicable in the context of one time passwords (e.g. initial passwords, activation codes). In addition we have established various new relations between the three notions of entropy (Shannon, Guessing, Min), providing strong improvements on existing bounds. Our results indicate that the expression $\log_2(2\alpha - 1)$, which we propose to call the *Searching* entropy, relates better to the other two entropies than the Guessing entropy α in its natural form.

It follows from Theorem 8 that distributions with fixed Guessing entropy α that satisfy $\log_2(2\alpha - 1) \leq H$ (an apparent popular belief) is of non-zero Lebesgue measure, i.e. the probability that a random distribution on n points with Guessing entropy equal to α satisfies this inequality is non-zero. It seems an interesting problem to establish the behavior of this probability in terms of α and n. A similar question is: what is the probability that a random distribution on n points satisfies $\log_2(2\alpha - 1) \leq H$? Based on our experiments it seems that this probability is close to one which we have actually shown for $n \leq 6$ as then all distributions satisfy this inequality.

8 Acknowledgments

Lisa Bloomer and Franklin Mendivil are thanked for discussing and eventually providing (Franklin) me with a technique for choosing random probability distributions on n points used in my simulations. Frans van Buul is thanked for writing the initial simulation software, Christian Cachin for discussions on the various types of entropy, Berry Schoenmakers for providing me some initial mathematical techniques and Marcel van de Vel for the discussions on convexity.

References

1. E. Arikan, *An inequality on guessing and its application to sequential decoding*, IEEE Trans. Inform. Theory, vol. 42, pp. 99-105, 1996.
2. A. Bosselaers, *Even faster hashing on the Pentium*, rump session presentation at Eurocrypt97, May 13, 1997.
3. C. Cachin, *Entropy Measures and Unconditional Security in Cryptography*, volume 1 of ETH Series in Information Security and Cryptography. Hartung-Gorre Verlag, Konstanz, Germany, 1997 (Reprint of Ph.D. dissertation No. 12187, ETH Zrich).

4. D.A. Huffman, *A method for the construction of minimum-redundancy codes*, Proceedings of the I.R.E., 1952, pp. 1098-1102
5. D. Malone, W.G. Sullivan, *Guesswork and entropy*, IEEE Transactions on Information Theory, Volume 50, Issue 3, pp. 525-526, 2004.
6. J.L. Massey, *Guessing and entropy*, Proc. 1994 IEEE International Symposium on Information Theory, 1994, p.204.
7. R.J. McEliece, Z. Yu, *An inequality on entropy*, Proc. 1995 IEEE International Symposium on Information Theory, 1995, p.329.
8. NIST, *Electronic Authentication Guideline*, Special Publication 800-63, 2004.
9. H.L. Royden, *Real analysis*, Macmillan Publishing company, New York, 1988.
10. Sci.crypt crypto FAQ, http://www.faqs.org/faqs/cryptography-faq/part04.
11. M. L. J. van de Vel, *Theory of Convex Structures*, North-Holland, 1993.

A Appendix: Notes on the model

Our model describes an ideal situation in which the computer system owner knows or can prescribe the probability distribution according to which users choose passwords. This is the case when the computer system owner generates the passwords for its users. In common practice, the system owner can at least highly influence the probability distribution by imposing "complexity rules" for passwords chosen by users.

In our model we have not taken the Shannon entropy as a measure for security of passwords which seems to be widely done. We have only found non-valid motivations for this, that are typically based on the misconception, e.g. in [8], [10], that the Shannon entropy and Guessing entropy are related by $\log_2(\alpha) = H$.

Perhaps contrary to popular belief, even an inequality of type

$$\log_2(2\alpha - 1) \leq H. \tag{9}$$

between the Shannon entropy and the Guessing entropy is not generally true as is shown by the earlier mentioned example of Massey [6].

In Theorem 8 we prove a variant on inequality (9) that does hold. We note that in this and many other results in this paper the expression $\log_2(2\alpha - 1)$ takes a prominent place. In fact, one can argue that this expression is a more suitable definition of Guessing entropy.

As Massey's bound can be rewritten as $\alpha \geq \log_2(H) - \log_2(e) \approx \log_2(H) - 1.4$ one can use the Shannon entropy in an underestimate for the Guessing entropy This indicates that the Shannon entropy can be used as an underestimate for resistance against complete off-line attacks. Without referring to Massey's bound, appendix A of [8] uses $\alpha \geq \log_2(H)$ and consequently from a theoretical perspective the numbers Table A.1 in [8] are about one bit too small. A large Shannon (and consequently Guessing) entropy does not provide resistance against incomplete attacks. To this end, for $0 < \delta < 1$ consider the distribution $(\delta, q_1, q_2, \ldots q_m)$ with $q_i = (1-\delta)/m$. This distribution has a Shannon entropy that goes to infinity when m goes to infinity, while the probability that the first guess is successful is δ irrespective of m. In other words, the Shannon and Guessing entropies alone are not an appropriate measure for secure password distribution as is suggested in table A.1 of [8].

In our model we have only considered an attacker that is after one specific password. In practice the attacker might test-data for several, say P, passwords (e.g. of different users), e.g. a UNIX "passwd" file or a Windows "SAM" database. If the attacker is only after one password (and not necessarily all of them or a specific one), his optimal strategy would be in parallel: trying if any of the passwords is the most likely password etcetera. If the test-data is just a secure hash of a password, then the attacker would be able to speed up the complete attack by about a factor P, which is why it is common practice to *salt* passwords before applying a secure hash function to it. If we assume that passwords are indeed adequately salted, the attacker does not gain any advantage from a parallel attack when he is aiming to mount a complete attack, i.e. finding all passwords. However the attacker gains advantage when he is aiming to mount an incomplete attack. Indeed if the attacker divides the computational effort he is willing to spend over the number of passwords, his probability of success is higher than if he only uses this effort to guess only one password. Our model can be used to quantify resistance against this attack as well by the following observation: if the probability of success of a incomplete attack with effort L against one password is q, then the probability of success of an incomplete P-parallel attack with effort $L \cdot P$ is $1-(1-q)^P \approx q \cdot P$.

B Appendix: proofs on convexity

Proof of Theorem 1: Note that C is a closed set. It suffices to prove the first part of the result as the remainder then follows from the Krein-Milman theorem. To this end, let $x \in C$ be an extreme point. Let be $S \subset \{1,...,s\}$ of highest cardinality such that

$$x \in \cap_{i=1}^{r} f^{-1}(\delta_i) \bigcap \cap_{j \in S} F^{-1}(\theta_j)$$

is of lowest affine dimension. Now suppose that this dimension is not zero, i.e. that x is not a minimal intersection point. Then first of all, $S \neq \{1,2,...,s\}$ as then set $\cap_{i=1}^{r} f^{-1}(\delta_i) \bigcap \cap_{j \in \{1,2,...,s\}} F^{-1}(\theta_j)$ would be an unbounded subset of C that is bounded by assumption. So $\{1,2,...,s\} \setminus S \neq \emptyset$ and for every $j \in \{1,2,...,s\} \setminus S$ it follows that $F_j(x) > \delta_j$. That is, there exists a small Euclidean ball B around x such that $F_j(y) > \delta_j$ for all $y \in B$. It now follows that

$$B \bigcap \cap_{i=1}^{r} f^{-1}(\delta_i) \bigcap \cap_{j \in \{1,2,...,s\}} F^{-1}(\theta_j) \subset C \tag{10}$$

Now, the intersection of a Euclidean ball and a affine space of dimension ≥ 1 contains more points than only its centre (i.e., x). Suppose that z is also in (10) than it easily follows that $x-(z-x) = 2x-z$ is also in (10). It then follows that $y = \frac{1}{2}(2y-z) + \frac{1}{2}z$ and x can not be an extreme point. We arrive at a contradiction and we conclude that x is a minimal intersection point.

Conversely, suppose that x is a minimal intersection point, i.e. there exists a S of $\{1, ..., s\}$, such that:

$$\{x\} = \cap_{i=1}^{r} f^{-1}(\delta_i) \bigcap \cap_{j \in S} F^{-1}(\theta_j). \tag{11}$$

Now suppose that x is not an extreme point, that is $x = \lambda y + (1-\lambda)z$ with $0 < \lambda < 1$ and $y, z \in C \setminus \{x\}$. It simply follows that for each $j \in S$ we have $F_j(y) = F_j(z) = \theta_j$ as otherwise $F_j(x) > \theta_j$. We conclude that y, z are also elements of the left hand side of (11) contradicting that x is minimal intersection point. $\square$

Proof of Theorem 2: The cases $n = 1$ and $n = 2$ can be easily verified directly and we assume that $n \geq 3$. Note that for $f_1(x) = \sum_{i=1}^{n} x_i$, $f_2(x) = \sum_{i=1}^{n} i \cdot x_i$, $F_j(x) = x_j - x_{j+1}$ for $1 \leq j < n$ and $F_n(x) = x_n$, the set $C_{n,\alpha}$ takes the form:

$$C_{n,\alpha} = \{x \in \mathbb{R}^n | f_1(x) = 1, f_2(x) = \alpha, \text{and } F_j(x) \geq 0 \text{ for } j = 1, 2..., n\}.$$

As $C_{n,\alpha}$ is clearly bounded we aim to use Theorem 1. For this we need to look for unique solutions of x of the equation $f_1(x) = 1, f_2(x) = \alpha$ and any subsets of the equations

$$F_1(x) = 0; F_2(x) = 0; ... F_n(x) = 0.$$

As any $n-3$ or smaller subset of these equations will certainly not result in unique solutions x, we only need to consider any $n-2$, $n-1$ and n-subset of these equations.

An $(n-2)$-subset pertains to leaving out two equations, say the j-th and the k-th with $1 \leq j < k \leq n$ which leads to the following system of equations:

$$\begin{aligned} x_1 = x_2 = ... &= x_j \\ x_{j+1} = x_{j+2} = ... &= x_k \\ x_{k+1} = x_{k+2} = ... &= x_n = 0. \end{aligned}$$

Together with the conditions $f_1(x) = 1, f_2(x) = \alpha$ simple calculations shows that this results in a unique solution $X_{n,j,k}$ as defined in the theorem. However these solutions also need to satisfy the remaining two equations, i.e., $x_j \geq x_{j+1}$ and $x_k \geq x_{k+1} = 0$. Simple calculations show that the first condition is equivalent to $k \geq 2\alpha - 1$ and that the second condition is equivalent to $j \leq 2\alpha - 1$. We conclude that the $X_{j,k,n}$ described in the theorem are all extreme points of $C_{n,\alpha}$.

An $(n-1)$-subset pertains to a $1 \leq j \leq n$ and

$$\begin{aligned} x_1 = x_2 = ... &= x_j \\ x_{j+1} = x_{j+2} = ... &= x_n = 0. \end{aligned}$$

If this has a solution, then $2\alpha - 1$ must be equal to j and hence an integer. The first j coordinates of this solution will be equal to $1/(2\alpha - 1)$ and the remaining ones are zero. We arrive at $X_{2\alpha-1,2\alpha-1,n}$. Finally, an n-subset cannot results in solutions, let alone unique ones. $\square$

Proof of Proposition 1: As the last equivalence is evident, we only prove the first one. To this end, let $(p_1, ..., p_n) \in C_{n,\alpha,d} \neq \emptyset$, then from Theorem 3 with $d = 1$ it follows that $1/(2\alpha - 1) \leq p_1$. If $(p_1, ..., p_n) \in C_{n,\alpha,d} \subset C_{n,\alpha}$ then clearly $p_1 \leq 1/d$. Hence it follows that $d \leq 2\alpha - 1$. The "only if" part of the first equivalence now follows from $C_{n,\alpha,d} \subset_{n,\alpha}$. Conversely suppose that $C_{n,\alpha} \neq \emptyset$ and $d \leq 2\alpha - 1$. Then according to Theorem 2 $X_{d,n,n}$ is one of the extreme points of Theorem 2 $C_{n,\alpha}$. It evidently follows that $X_{d,n,n} \in C_{n,\alpha,d}$, showing that this set is not empty. Actually, this alternatively follows from $X_{\lfloor 2\alpha-1 \rfloor,n,n} \in C_{n,\alpha,d}$. □

Proof of Proposition 2: We start the proof with two observations. First, from the description of the points $X_{j,k,n}$ in Theorem 2 it follows that all first j probabilities are strictly larger than the remaining ones provided that $k \neq 2\alpha - 1$. Second, if $k = 2\alpha - 1$ (hence $2\alpha - 1$ is an integer in particular), then for all $1 \leq j \leq 2\alpha - 1$, the points $X_{j,k,n}$ are equal to the distribution consisting of $2\alpha - 1$ non-zero probabilities equal to $1/(2\alpha - 1)$. As $d \leq 2\alpha - 1$ by Proposition 1 it evidently follows that then all first d probabilities are equal in $X_{j,k,n}$.

For a proof of the proposition; by using the description of the points $X_{j,k,n}$ in Theorem 2 it directly follows that if $d \leq j \leq 2\alpha - 1 \leq k$ that then the first d probabilities are equal.

Conversely, let a point $Y \in C_{n,\alpha}$ have its first d probabilities equal. Then by Theorem 2 the point Y is the convex combination of points $X_{j,k,n}$ satisfying $1 \leq j \leq 2\alpha - 1 \leq k$, $j < k$. Suppose that in this convex combination, some points $X_{j,k,n}$ contribute that do not satisfy $d \leq j$, i.e. $d > j$. If for some of these points $k = 2\alpha - 1$ then, by the second observation at the beginning of the proof, the contribution of this $X_{j,k,n}$ can be replaced with $X_{d,k,n}$.

So we may assume for the point $X_{j,k,n}$ contributing to Y satisfies $d > j$ and $k > 2\alpha - 1$. Let j' be the smallest j with this condition, it follows from the first observation at the beginning of the proof, that the first $j' < d$ probabilities in Y are strictly larger than the $j' + 1$-th probability and hence that the first d probabilities in Y are not equal. We arrive at a contradiction. □

Proof of Theorem 6: We number the points in E, i.e. $E = \{e_1, e_2, \ldots, e_{|E|}\}$. Clearly,

$$C_{n,\alpha,\delta} = \{\sum_{i=1}^{|E|} \lambda_i e_i \mid e_i \in E, \sum_{i=1}^{|E|} \lambda_i = 1, \lambda_i \geq 0, \sum_{i=1}^{|E|} \lambda_i (e_i)_1 \leq \delta\}. \tag{12}$$

We relate $C_{n,\alpha,\delta}$ with

$$L = \{(\lambda_1, \lambda_2, \ldots, \lambda_{|E|}) \mid \lambda_i \geq 0, \sum_{i=1}^{|E|} \lambda_i = 1, \sum_{i=1}^{|E|} \lambda_i (e_i)_1 \leq \delta\} \subset \mathbb{R}^{|E|}.$$

For $l = (\lambda_1, \lambda_2, \ldots, \lambda_{|E|}) \in L$ we define $l \cdot E = \sum_{i=1}^{|E|} \lambda_i \cdot e_i$. As the set L is convex, closed and bounded it is spanned by it extreme points. Following Theorem 1 the extreme points of L are of type

$$F = \{(\lambda_1, \lambda_2, \ldots, \lambda_{|E|}) \mid \quad \text{only two different } \lambda_i, \lambda_j \in [0, 1] \text{ are non-zero and}$$

$$\lambda_i + \lambda_j = 1, \lambda_i \cdot (e_i)_1 + \lambda_j \cdot (e_j)_1 = \delta$$
$$\text{or } \lambda_i = 1 \text{ and } (e_i)_1 \leq \delta\}$$

Clearly, F is finite. Also, any convex combination $(\lambda_1, \lambda_2, \ldots, \lambda_{|E|})$ occurring in (12) is a convex combination of elements in F. In other words, the convex hull of $F \cdot E$ is equal to $C_{n,\alpha,\delta}$. That is, an extreme point of $C_{n,\alpha,\delta}$ is either an extreme point f in $C_{n,\alpha}$ with $(f)_1 \leq \delta$ or of type

$$\lambda X_{j_1,k_1,n} + (1-\lambda) X_{j_2,k_2,n} \tag{13}$$

with $1 \leq j_1, j_2 \leq 2\alpha - 1 \leq k_1, k_2 \leq n$ and $\lambda \in (0,1)$.

We now take another view at the extreme points of $C_{n,\alpha,\delta}$. Using the technique used in the proof of Theorem 2 it follows that the extreme points of $C_{n,\alpha,\delta}$ are either $f \in E$ with $(f)_1 \leq \delta$ or take the form

$$\begin{array}{cccccccccc} (\ \delta, & \delta, & \cdots \delta, & a, & \cdots a, & b, & \cdots b, & 0, & \cdots 0) \\ & & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\ 1, & 2, & \cdots j, & j+1, & \cdots k, & k+1, & \cdots m, & m+1 & \cdots n \end{array} \tag{14}$$

with the condition that $\delta \geq a \geq b > 0$ and that this point is in $C_{n,\alpha}$. [2]

If either k_1 or k_2, say k_1, in expression (13) is equal to $2\alpha - 1$ (also implying that $2\alpha - 1$ is an integer) then this expression also holds if we take $j_1 = j_2$. Indeed, all points of type $X_{j,2\alpha-1,n}$ are equal (cf. Theorem 2). So assume that $2\alpha - 1 < k_1, k_2$. As is shown by Theorem 2 all extreme points $X_{j,k,n}$ with $2\alpha - 1 < k$ in E are of a special form: the first j coordinates are equal and strictly larger than the $j+1$ to k-th coordinates which are also equal and strictly larger than zero. It follows immediately that a point as in expression (13) can only be of the prescribed form (14) if either $j_1 = j_2$ or $k_1 = k_2$. □

C Appendix: proof of Theorem 8

Lemma 3 *Let α, k, m, n with $1 \leq 2\alpha - 1 \leq k \leq m \leq n$. Then the function $(0, 2\alpha - 1] \ni j \to \min(G(j, 2\alpha - 1, k), \log_2(2\alpha - 1))$ is increasing.*

Proof: The function $F_k : (0, 2\alpha - 1] \ni j \to G(j, 2\alpha - 1, k)$ is concave by Lemma 2 so for any $j \in (0, 2\alpha - 1]$:

$$\begin{aligned} \min\{F_k(x) \mid x \in [j, 2\alpha - 1]\} &= \min(F_k(j), F_k(2\alpha - 1)) \\ &= \min(F_k(j), \log_2(2\alpha - 1)) \\ &= \min(G(j, 2\alpha - 1, k), \log_2(2\alpha - 1)). \end{aligned}$$

As the minimum of a function on the descending interval $[j, 2\alpha - 1]$, the function $[0, 2\alpha - 1] \ni j \to \min(G(j, 2\alpha - 1, k), \log_2(2\alpha - 1))$ is increasing. □

[2] We note that, unlike in the proof of Theorem 2, not all points of the prescribed form (14), automatically satisfy the remaining conditions.

Lemma 4 *Let α, k, m, n with $1 \le 2\alpha - 1 \le k \le m \le n$ and let $\delta \ge 1/(2\alpha-1)$, then*

$$\min(G(g_k^\alpha(\delta), 2\alpha-1, k), \log_2(2\alpha-1)) \ge \min(G(g_m^\alpha(\delta), 2\alpha-1, m), \log_2(2\alpha-1)).$$

Proof: It is easily verified that for any fixed $x \ge 1/(2\alpha-1)$ the map $[k,\infty) \ni l \to g_l(x)$ is increasing in l. Hence we have $g_k^\alpha(\delta) \le g_m(\delta)$ and it follows from Lemma 3 that

$$\min(G(g_k^\alpha(\delta), 2\alpha-1, k), \log_2(2\alpha-1)) \ge \min(G(g_m^\alpha(\delta), 2\alpha-1, k), \log_2(2\alpha-1)).$$

From the third part of Lemma 2 it follows that the function $[2\alpha-1,\infty) \ni z \to \min(G(j, 2\alpha-1, z), \log_2(2\alpha-1))$ is decreasing. We conclude that

$$\min(G(g_k^\alpha(\delta), 2\alpha-1, k), \log_2(2\alpha-1)) \ge \min(G(g_m^\alpha(\delta), 2\alpha-1, m), \log_2(2\alpha-1)),$$

finishing the proof of the lemma. □

Lemma 5 *Let α be fixed, $1 \le 2\alpha-1 \le n$, and let $\delta \ge 1/(2\alpha-1)$ then the sequence $\{\min(G(g_m(\delta), 2\alpha-1, m), \log_2(2\alpha-1))\}_{m \ge n}$ converges to $-\log_2(\delta)$.*

Proof: By Lemma 1 it follows that

$$G(g_m(\delta), 2\alpha-1, m) = \log_2(g_m(\delta)) + G(1, (2\alpha-1)/g_m(\delta), m/g_m(\delta)).$$

Observe that $\lim_{m\to\infty} g_m(\delta) = 1/\delta$ and

$$\lim_{m\to\infty} G(1, (2\alpha-1)/g_m(\delta), n/g_m(\delta)) = \lim_{k\to\infty} G(1, (2\alpha-1)\delta, k) = 0.$$

Hence

$$\lim_{n\to\infty} G(g_n(\delta), 2\alpha-1, n) = -\log_2(\delta).$$

The lemma now follows from the fact that $\delta \ge 1/(2\alpha-1)$. □

Lemma 6 *Let $1 \le 2\alpha-1 \le k$, $1/(2\alpha-1) \le x \le 1 - 2(\alpha-1)/k$ and let $\alpha' = (\alpha-1)/(1-\delta)$ then*

1. $2(\alpha-1) \le 2\alpha'-1 \le k-1$
2. $(1-x)\log_2(2\alpha'-1) + h(x) = G(1, 2\alpha-1, 2\alpha')$.
3. $G(g_{k-1}^{\alpha'}(x/(1-x)), 2\alpha'-1, k-1) = (G(g_k^\alpha(x), 2\alpha-1, k) - h(x))/(1-x)$.

Proof: The first part of the lemma is a straightforward verification. The second part of the lemma follows from the second part of Lemma 1 by taking $y = 2\alpha-1$ and $z = 2(\alpha-1)/(1-x) = 2\alpha'$.

For a proof of the last part of the lemma, consider the following series of equalities for the expression $-G(g_k^\alpha(x), 2\alpha-1, k) + h(x)$.

$$\begin{aligned}
&= g_k^\alpha(x) x \log_2(x) + (1 - g_k(x)x)\log_2((1 - g_k^\alpha(x)x)/(k - g_k^\alpha(x))) \\
&\quad -x\log_2(x) - (1-x)\log_2(x) \\
&= ((g_k^\alpha(x) - 1)x\log(x) - (1-x)\log_2(x) + \\
&\quad +(1 - g_k^\alpha(x)x)\log_2((1 - g_k^\alpha(x)x)/(k - g_k^\alpha(x)))
\end{aligned}$$

Placing the equality $(1-x)\log_2(1-x) = (g_k^\alpha(x)-1)x\log_2(1-x) + (1-g_k^\alpha(x)x)\log_2(1-x)$ in the last expression yields:

$$= ((g_k^\alpha(x)-1)x\log(x/(1-x)) + (1-g_k^\alpha(x)x)\log_2(\frac{(1-g_k^\alpha(x)x)}{(k-g_k^\alpha(x))(1-x)})$$

Now as it is simply verified that $g_k^\alpha(x)-1 = g_{k-1}^{\alpha'}(x/(1-x))$ the last expression is equal to

$$\begin{aligned}
&(g_{k-1}^{\alpha'}(x/(1-x))x\log_2(x/(1-x)) + \\
&(1-x-g_{k-1}^{\alpha'}(x/(1-x))x\log_2(\frac{(1-x-g_{k-1}^{\alpha'}(x/(1-x))x}{(k-1-g_{k-1}^{\alpha'}(x/(1-x)))(1-x)}) \\
&= g_{k-1}^{\alpha'}(x/(1-x))x\log(x/(1-x)) + \\
&(1-g_{k-1}^{\alpha'}(x/(1-x)x/(1-x)))\log_2(\frac{(1-g_{k-1}^{\alpha'}(x/(1-x)x/(1-x))}{k-1-g_{k-1}^{\alpha'}(x/(1-x))} \\
&= (1-x)G(g_{k-1}^{\alpha'}(x/(1-x)), \frac{2(\alpha-1)}{1-x}-1, k-1).
\end{aligned}$$

The last part of the lemma is now immediate.

□

Lemma 7 *Let $\bar{p} = (p_1,\ldots,p_n) \in C_{n,\alpha,\delta}$ with $p_1 = \delta < 1$. Then*

1. $\bar{p'} = (p_2,\ldots,p_n)/(1-\delta) \in C_{n-1,(\alpha-1)/(1-\delta),\delta/(1-\delta)}$
2. $H(\bar{p'}) = \frac{H(\bar{p})-h(\delta)}{1-\delta}$.

Proof: This is a straightforward verification. □

Proof of Theorem 8: If δ is not admissible (i.e. does not satisfy equality (6)) then $g_n^\alpha(\delta) \leq 1$ and $C_{n,\alpha,\delta} = C_{n,\alpha}$. It follows from Theorem 4 and Lemma 3 that

$$\begin{aligned}
H(n,\alpha,\delta) = M_{n,\alpha,1} &\geq \min(G(1, 2\alpha-1, n), \log_2(2\alpha-1)) \\
&\geq \min(G(g_n^\alpha(\delta), 2\alpha-1, n), \log_2(2\alpha-1)).
\end{aligned}$$

From now on we may assume that δ is admissible. By concavity of the Shannon entropy it suffices to show that the entropy on points in the set E' defined in Theorem 6 take values greater or equal than stated in the first part of the result. From Theorem 6 it follows that two types of points $p \in E'$ exist:

1. points p of the form $X_{j,k,n}$ with $1 \leq j \leq 2\alpha-1 \leq k \leq n$ such that $a_{j,k,n} \leq \delta$,
2. points p of the form $\lambda X_{j_1,k_1,n} + (1-\lambda)X_{j_2,k_2,n}$ with $1 \leq j_1, j_2 \leq 2\alpha-1 \leq k_1, k_2 \leq n$ such that $\lambda a_{j_1,k_1,n} + (1-\lambda)a_{j_2,k_2,n} = \delta$.

With respect to the first case; as the function $g_k^\alpha(\cdot)$ is decreasing, it follows from $a_{j,k,n} \leq \delta$ that $j = g_k^\alpha(a_{j,k,n}) \geq g_k^\alpha(\delta)$.

$$\begin{aligned}
H(X_{j,k,n}) &= G(j, 2\alpha - 1, k) \geq \min(G(j, 2\alpha - 1, k), \log_2(2\alpha - 1)) \\
&\geq \min(G(g_k^\alpha(\delta), 2\alpha - 1, k), \log_2(2\alpha - 1)) \\
&\geq \min(G(g_n^\alpha(\delta), 2\alpha - 1, n), \log_2(2\alpha - 1)).
\end{aligned}$$

where the second inequality follows from Lemma 3 and the last inequality is Lemma 4.

With respect to the remaining second case; we proceed by using induction to n. Note that theorem is clearly true for $n = 1, 2, ..., 6$ as we have in fact shown that inequality (9) holds for these n. So suppose that the theorem holds for $n-1$. Write $p = (p_1, \ldots, p_n)$ with $p_1 = \delta$ and consider $p' = (p_2, \ldots, p_n)/(1 - \delta)$ and $\alpha' = (\alpha - 1)/(1 - \delta)$. Then according to Lemma 7 $p' \in C_{n-1,(\alpha-1)/(1-\delta),\delta/(1-\delta)} \neq \emptyset$ and $H(p') = \frac{H(p)-h(\delta)}{1-\delta}$. It now follows that

$$\begin{aligned}
H(p) &= (1 - \delta) * H(p') + h(\delta) \\
&\geq (1 - \delta) \min \Big(G(g_{k-1}^{\alpha'}(x/(1 - x)), 2\alpha' - 1, k - 1), \\
&\qquad\qquad \log_2(2\alpha' - 1)) + h(\delta) \\
&= \min(G(g_k^\alpha(x), 2\alpha - 1, k), G(1, 2\alpha - 1, 2\alpha')),
\end{aligned}$$

where the first inequality is the inductive assumption and the last equality is Lemma 6. To finish the induction we need to prove that

$$G(1, 2\alpha - 1, \frac{2(\alpha - 1)}{1 - \delta}) \geq \min(G(g_k^\alpha(\delta), 2\alpha - 1, k), \log_2(2\alpha - 1)).$$

To this end, if we fix α then the first term occurring in the inequality above only depends on δ, the second term depends only on k and δ and the last term is constant. Denote the terms in the inequality by $A(\delta)$, $B(k, \delta)$ and C respectively and we need to prove that $A(\delta) \geq \min(B(k, \delta), C)$. If $A(\delta) \geq C$ we are done, so suppose that $A(\delta) \leq C$.

As δ is admissible it follows that $1/(2\alpha - 1) \leq \delta \leq 1 - 2(\alpha - 1)/k$. This implies that there exists a $2\alpha - 1 \leq k' \leq k$ such that $\delta = 1 - 2(\alpha - 1)/k'$. It is easily verified that $A(\delta) = B(k', \delta) = G(1, 2\alpha - 1, k')$. By Lemma 4 it now follows that

$$\begin{aligned}
\min(B(k, \delta), C) &\leq \min(G(g_{k'}^{\alpha'}(\delta), 2\alpha - 1, k'), \log_2(2\alpha - 1)) \\
&= \min(A(\delta), \log_2(2\alpha - 1)) = A(\delta).
\end{aligned}$$

□

D Appendix: comparison of bounds

In Figure 1 below we have for $n = 23$ and $\alpha = 7$ depicted the graphs of $\delta \to H(n, \alpha, \delta)$ calculated using Theorem 7 labeled by "A"; $\delta \to \min(G(g_n^\alpha(\delta), 2\alpha -$

$1, n), \log_2(2\alpha - 1))$ labeled by "B", $\delta \to G(g_n^\alpha(\delta), 2\alpha - 1, n)$ labeled by "C", the bound in Theorem 9 labeled by "D" and the Min entropy $\delta \to -\log_2(\delta)$ labeled by "E". Finally we have depicted the 13 points $(a_{j,n,n}, H(X_{j,n,n}))$. It is easily verified that Theorem 8 is strong in the sense that for all δ that equal the first coordinate of an extreme point of type $X_{j,n,n}$, i.e. $\delta = a_{j,n,n}$ equality in Theorem 8 holds provided $H(X_{j,n,n}) \le \log_2(2\alpha - 1)$. However, the figure below indicates that these distributions are actually "local" minima with respect to the Shannon entropy. The figure also indicates that the bound in Theorem 8 is strong, certainly in comparison with the bound in Theorem 9 and the Min entropy. We finally note that the example also shows that taking the minimum with $\log_2(2\alpha - 1)$ in Theorem 8 cannot be relaxed.

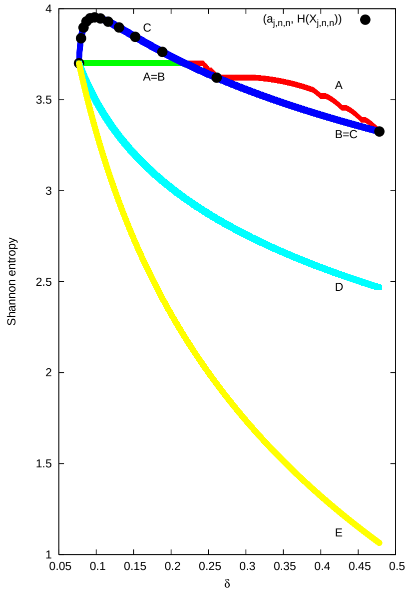

Fig. 1. Comparison of bounds