A MEASURE OF SEMIEQUIVOCATION

Andrea Sgarro

Department of Mathematics and Computer Science
University of Udine
33100 Udine, Italy

ABSTRACT

A Shannon-theoretic cryptographic model is described in which the purpose of the cryptanalist is to find a set of M elements containing the solution, rather than finding the solution itself. For $M = 2$ we introduce the notions of semientropy, semiequivocation and duplicity distance, which are counterparts to well-known notions met in the case $M = 1$. It is argued that in some situation our model takes into account the semantical competence of the cryptanalist.(as opposed to his statistical competence) better than the usual model does.

I. Introduction

In Shannon-theoretic cryptography the clearmessage source is usually described as a stochastic process. In the literature results have appeared for substitution and transposition ciphers (cf e.g. /1/ to /5/) which hold assuming that the message source has a well-defined statistical behaviour, for example that it is memoryless and stationary; the letter probabilities might be given, as in /1/ to /5/, or might be left unspecified. The latter point of view is called "universal" in non-secret coding theory, but we feel that this term is rather misleading in the context to follow (statisticians prefer the less ambitious term "robust"). In /6/ also the case of Markov sources is covered.

Let us assume that the clearmessage is written in a natural language like English. Describing English as a Markov process with memory 3 is often considered to be reasonably adequate; actually, in non-secret coding much coarser descriptions have brought forth considerable practical

C.G. Guenther (Ed.): Advances in Cryptology - EUROCRYPT '88, LNCS 330, pp. 375-387, 1988.

success, starting with the Morse code of 1838. As a matter of fact, a natural language results from the superposition of comparatively simple frequency-type dependences and extremely complicated semantical dependences which can act even on a very large range. In principle, also this latter type of dependences can be captured in a single all-comprehensive statistical description: its intricacy, however, is far past the possibility of numerical assessments.

In cryptography, unlike in non-secret coding, keeping only short-range frequency-type dependences does not seem to be a wise policy. Frequency-type descriptions are too optimistic, because they ignore the semantical competence of the cryptanalyst. This is the opposite of what one should do in cryptography, where, if need be, models have to be over-pessimistic, and not the other way round.

In /6/ and /7/ this author has pointed out certain unpleasant "paradoxes" which result from assuming that the clearmessage source has a simple and well-defined statistical behaviour, like, say, memoryless and stationary, or Markov with given memory. Certainly, no paradox arises in the case of results which are "universal" in the proper sense of this term. Take the perfection of the one-time pad (cf e.g. /8/), which holds true whatever the message statistics may be; no assumption is needed, not even ergodicity or stationarity. This result, covering even the most misterious long-range dependences, is perfectly sound and ready to be used. Unfortunately, accepting only results which have such universal validity is a very restrictive policy, indeed.

Below we describe an alternative Shannon-theoretic model which takes inspiration from historical cryptanalytic practice. The idea is the following: a cryptanalyst would first use his frequency-type statistical knowledge to curtail the number of possible solutions; when this number is small, semantics gets the upper hand of frequency-type statistics, and he can find directly the solution without further bother. In other words, the purpose of the spy is not to find the solution by frequency-type arguments, but only to find a "small" set of possible solutions. In the following we shall fix an integer M and declare "small" a set with M elements; actually in calculations we shall go so far as to take $M = 2$. Of course, this is quite arbitrary; however, our purpose here is to explore the quantitative variations in the new case $M = 2$ with respect to the classical case $M = 1$ to derive qualitative information for the more general situation $M > 1$. Observe that since we assume that the first part of the spy's job is statistical "stricto sensu" we are justified in using those

neat descriptions for the behaviour of the message source which we have argued to be fishy in the case $M = 1$.

Our approach leads us to define a new measure of equivocation, which we call *semiequivocation*. Key equivocation, say, represents the uncertainty of the spy who has intercepted the cryptogram and wants to identify the correct key (cf e.g. /8/); instead, key semiequivocation will represent the uncertainty of the spy who only wants to find a doubleton containing the correct key. Equivocation is a conditional entropy; its meaning is based on the fact that Shannon's entropy is an adequate measure of statistical uncertainty. Before introducing our new measure of semiequivocation, we shall have to introduce an (unconditional) new measure of "semi-uncertainty", called *semientropy*, which will be the counterpart to Shannon's entropy. This will be done in section II, while section III is devoted to the notions of semiequivocation and *duplicity distance*, the latter being the counterpart to that of unicity distance (cf e.g. /8/); an example is given. Section IV contains a final comment.

We adopt the notation of /9/ for information-theoretic concepts; in particular, $H(X) = H(P)$ is the entropy of the random variable (r.v.) X with probability distribution (p.d.), or probability vector, $P = (p_1, p_2, \ldots, p_K)$, while $I(X;Y) = I(P,W)$ is the mutual information between the r.v.'s X and Y, the probability distribution of this random couple being determined by the p.d. P of X and the stochastic matrix W which gives the conditional probabilities of Y given X; $h(p)$ is the binary entropy function: $h(p) = H(P)$ with $P = (p, 1-p)$; $D(P \mid Q)$ is the informational divergence (cross-entropy) of P and Q, in this order. Logarithms are to any base greater than 1. The source alphabet is $\aleph = \{a_1, a_2, \ldots, a_K\}, K \geq 2$; we shall write indifferently p_i or $P(a_i)$.

II. Semientropy

Shannon's entropy is considered to be an adequate measure of statistical uncertainty. There are several justifications, both "axiomatic" and "pragmatic", to this interpretation (cf e.g. /9/). The pragmatic point of view derives the meaning of entropy from coding theorems which, roughly speaking, state that $H(X) = H(P)$ is the minimum (not necessarily integer) number of bits needed to reliably describe the outcome of r.v. X; these bits are "nearly" independent and equidistributed, so that each binary digit contains almost exactly one binary bit of information (cf /9/);

(we suppose here that the logs are to the base 2). In our setting a "reliable description" of the outcome of X must be understood in a slacker sense. Actually, we are not interested in knowing the exact value of X, but rather in finding out an M-set to which this value belongs. We shall take inspiration from rate-distortion theory. Let us take a reproduction alphabet whose "letters" are the M-sets of primary letters (we assume $M \leq K$); let us consider a distortion measure $d(a, y)$ which is zero iff (if and only if) letter a belongs to set y (one may define $d(a, y) = 1$ otherwise, but this is irrelevant for zero distortions). We shall resort to $R_P(0)$, the rate-distortion function computed for distortion level 0, to measure the "reduced uncertainty" contained in X which is relevant to us. Of course, for $M = 1$ one re-finds Shannon's entropy; for $M = 2$, $R_P(0)$ will be called the semientropy of X, or of P, and denoted by $S(X) = S(P)$. In the following, unless otherwise specified, we assume $M = 2$.

$S(P)$ represents the minimum (not necessarily integer) number of bits (of D-its if logs to the base D are used) needed to reliably describe the outcome of X, taking into account our reduced needs of fidelity with respect to the classical case $M = 1$.

Definition 1. The *semientropy* $S(X) = S(P)$ of r.v. X with p.d. P is defined as

$$S(X) = S(P) = \min_{W} I(P, W) = \min_{XY} I(X; Y)$$

Above the first minimum is taken with respect to stochastic matrices W such that $W(y \mid a) > 0$ implies $a \in y$, or $d(a, y) = 0$; the second with respect to a random couple XY with distribution given by P and W, W as above. Corollary 2.3.7. in /9/ allows us to give an alternative definition of $S(P)$:

$$S(P) = -\min_{Q} \{D(P \mid Q) + \max_{i \neq j} \log[q_i + q_j]\} \tag{1}$$

Above $Q = (q_1, q_2, \ldots, q_K)$ is a d.p. over the primary alphabet $\aleph$.

The theorem below gives an explicit formula for $S(P)$.

Theorem 1.

$$S(P) = H(P) - \log 2 \quad \text{if} \quad p^* \leq \tfrac{1}{2},$$
$$S(P) = H(P) - h(p^*) \quad \text{if} \quad p^* \geq \tfrac{1}{2},$$

p^* being the largest probability in P.

Proof. The bound $S(P) \geq H(P) - \log 2$ follows from $I(X;Y) = H(X) - H(X \mid Y) = H(P) - H(X \mid Y)$ because, given doubleton Y, X takes at most two values with positive probability, the two elements of Y. We explore the conditions for equality in that bound. The bound is attained when an admissible W exists such that $P(a \mid y) = P(b \mid y) = \frac{1}{2}$ for any $y = \{a, b\}$ such that $R(y) > 0$ (the notation is self-explaning; R is the marginal distribution over the secondary alphabet). Therefore the criterion for having equality in the bound is $W(y \mid a) = \frac{R(y)}{2P(a)}$, $a \in y$. Suppose R has been fixed over the set of couples. A W giving that R exists iff, for each a:

$$\sum_{y:a\in y} \frac{R(y)}{2P(a)} = 1 \left[= \sum_{y:a\in y} W(a \mid y)\right]$$

(non-negativity for W is ensured by non-negativity for P). Therefore the lower bound is attained iff the system:

$$\text{for all } a \quad \sum_{y:a\in y} R(y) = 2P(a) \tag{2}$$

has non-negative solutions $R(y)$ (these sum to 1 as ensured by $\sum p_i = 1$: the sum of the first sides is $2\sum_y R(y)$). For $p^* > \frac{1}{2}$ the system is clearly impossible. In the Appendix we prove that the system does admit of positive solutions for $p^* \leq \frac{1}{2}$. Then $S(P) = H(P) - \log 2$ for $p^* \leq \frac{1}{2}$, $S(P) > H(P) - \log 2$ for $p^* > \frac{1}{2}$. Fix letter a and use a test matrix W defined as follows ($a \neq b$):

$$W(\{a,b\} \mid a) = \frac{P(b)}{1 - P(a)}, \qquad W(\{a,b\} \mid b) = 1, \quad \text{else zero.}$$

A computation shows that in this case $I(P,W) = H(P) - h(P(a))$. Then, for all i : $S(P) \leq H(P) - h(p_i)$. Consider now the alternative definition (1) of $S(P)$. Without real restriction assume $p_1 = p^*$; we shall use the test distribution Q with components proportional to $(p^*, 1-p^*, 1-p^*, \ldots, 1-p^*)$. If $p^* \geq \frac{1}{2}$, or $p^* \geq 1 - p^*$, one obtains after a few calculations:

$$-\{D(P \mid Q) + \max_{a\neq b} \log[Q(a) + Q(b)]\} = H(P) - h(p^*)$$

Therefore, for $p^* \geq \frac{1}{2}$: $S(P) \geq H(P) - h(p^*)$. Combining the two inequalities for $S(P)$ one has $S(P) = H(P) - h(p^*)$ for $p^* \geq \frac{1}{2}$. QED

As a corollary to the theorem we soon obtain a list of properties of $S(P)$ which vindicate its interpretation as an uncertainty measure to be used when the "experimenter" does not care about the precise value taken by r.v. X, but is satisfied as soon as he knows a doubleton containing X:

Corollary 1.

i) $S(P)$ is a concave function of P;

ii) $0 \leq S(P) \leq \log \frac{K}{2}$; $S(P) = 0$ iff P has at most two positive components; for $K > 2$: $S(P) = \log \frac{K}{2}$ iff P is uniform;

iii) $H(P) - \log 2 \leq S(P) \leq H(P)$; $S(P) = H(P) - \log 2$ iff $p^* \leq \frac{1}{2}$ (cf theorem 1); $S(P) = H(P)$ iff $H(P) = 0$, that is iff X is deterministic.

The properties in i) and ii) are obvious counterparts to similar properties of Shannon's entropy $H(P)$; we stress that, as soon as there are at least three positive probability letters, $S(P)$ is positive too. In iii) $S(P)$ and $H(P)$ are compared; the inequality $S(P) \leq H(P)$ is always strict in the non-deterministic case. The difference $H(P) - S(P)$ is largest when the uncertainty $H(P)$ is "large", in the sense there is no single "event" of "high" probability.

Remark. Observe that similar properties with M instead of 2 can be derived also in the general case $2 \leq M \leq K$ directly from the definition of $S(P)$ extended to the case $M > 2$ (take the secondary alphabet to be the set of M-sets of primary letters; in the alternative definition (1) one has to consider the sum of the M, and not of the two, most Q-probable letters). Property i) is a general property of the rate-distortion function for fixed distortion-level and follows from the (weak) concavity in P of $I(P, W)$ (cf /9/); the left side of ii) is trivial; the right side can be obtained from representation (1) computing the maximum in P of the right side of (1) and interchanging the two extrema; the left side of iii) can be obtained generalizing the arguments given at the beginning of the proof of the theorem; the right side is trivial. We go back to the case $M = 2$.

The concavity of $S(P)$ is not strict, since $S(P) = 0$ for all P with at most two positive components. Theorem 2 below deepens property i). It turns out that there is more linearity than that brought about by the case $S(P) = 0$; therefore, from the point of view of concavity $S(P)$ and $H(P)$ exhibit an important difference of behaviour (cf the discussion after corollary 2 below).

Theorem 2. Consider the closed segments of p.d.'s of the following form:

i) $[R, Q]$, with R and Q deterministic, $R \neq Q$;

ii) $[R, Q]$, with $r_i = \frac{1}{2}, q_i = 1$.

$S(P)$ is linear over all segments of this form and nowhere else. If P is a p.d. over a segment of type i) one has $S(P) = 0$; if P is a p.d. over a segment of type ii) one has $S(P) = 2(1 - p_i)S(R) = 2(1 - p_i)[H(R) - \log 2]$.

Proof. In the "inner region" $\max p_i \leq \frac{1}{2}$, $S(P)$ is strictly concave, $H(P)$ being so. Let us go to the "outer region" $\max p_i \geq \frac{1}{2}$ (the regions' frontiers overlap). The case i) when $S(P) = 0$ has already been disposed of. We go to case ii) assuming $K \geq 3$ else $S(P)$ is identically zero. A p.d. P over $[R, Q]$ has the form $P = (p_1, \varrho r_2, \varrho r_3, \ldots, \varrho r_K)$, $\frac{1}{2} \leq p_1 \leq 1$, $\varrho = 2(1 - p_1)$, $0 \leq \varrho \leq 1$ (we have taken $i = 1$ without real restriction). A computation shows that:

$$S(P) = H(P) - h(p_1) = \varrho[H(R) - \log 2] = \varrho S(R) \tag{3}$$

Clearly, $S(P)$ cannot be linear over a proper super-segment of $[R, Q]$, else one would trespass into the inner region. We have still to prove that $S(P)$ is linear only over segments i) and ii). Take R and Q distinct in the outer region. First assume that R and Q have their maximum in the same position, say the first. Then this is true also of the outer region point $V = \frac{1}{2}R + \frac{1}{2}Q$. Assume $S(P)$ is linear over segment $[R, Q]$. Then $S(V)$ can be computed in two ways (use linearity and (3)):

$$S(V) = \frac{1}{2}S(R) + \frac{1}{2}S(Q) = (1 - r_1)S(\tilde{R}) + (1 - q_1)S(\tilde{Q})$$

and

$$S(V) = 2(1 - v_1)S(\tilde{V})$$

Above $\tilde{R}$, $\tilde{Q}$, $\tilde{V}$ are suitable p.d.'s over the region intersection with $\tilde{r}_1 = \tilde{q}_1 = \tilde{v}_1 = \frac{1}{2}$. By comparison, recalling that $v_1 = \frac{1}{2}r_1 + \frac{1}{2}q_1$:

$$(1 - r_1)H(\tilde{R}) + (1 - q_1)H(\tilde{Q}) = (2 - r_1 - q_1)H(\tilde{V})$$

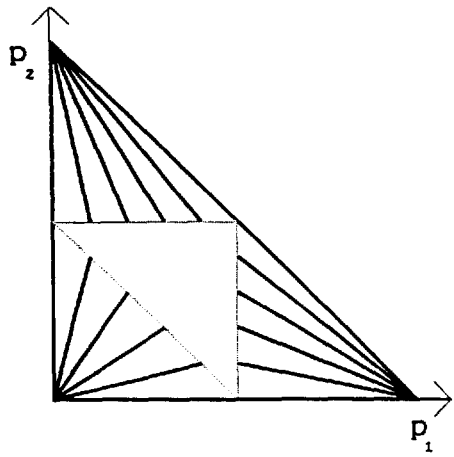
or:

$$\sigma H(\tilde{R}) + (1 - \sigma)H(\tilde{Q}) = H(\tilde{V}), \quad \text{with} \quad \sigma = \tfrac{1 - r_1}{2 - r_1 - q_1}$$

(the denominator is not zero, because R and Q are distinct). Actually, $\tilde{V} = \sigma\tilde{R} + (1 - \sigma)\tilde{Q}$, as a computation shows (convert the definition of V

into an equality for $\tilde{V}$, $\tilde{R}$ and $\tilde{Q}$). It is enough to observe that $H(P)$ is strictly concave to conclude $\tilde{V} = \tilde{R} = \tilde{Q}$; then V, R and Q lie on one of the old segments. Assume now that R has its maximum in the first position, while Q in the second, say. If $r_1 = q_2 = \frac{1}{2}$, the open segment $]R, Q[$ lies in the inner region, and there $S(P)$ is strictly concave. If $r_1 > \frac{1}{2}$, say, there is a sub-segment of $[R, Q]$ with positive length for whose points the first component is at least $\frac{1}{2}$. Taking into account this sub-segment, we go back to the cases already dealt with. QED

The figure shows some of the linearity segments in the case $K = 3$; the dotted lines show the region intersection.



III. Semiequivocation and duplicity distance

Below we deal with the case $M = 2$; however, much of what follows can be extended to the case of any M (cf the remark in section II).

So far we have defined a measure of unconditional "semi-uncertainty". Now we define a measure of conditional semi-uncertainty. Assume XC is a finite random couple; for convenience X will be interpreted as the random key (also the random message would be a suitable interpretation) and C as the random cryptogram. For an observed cryptogram c, $S(X \mid C = c)$, the unconditional semientropy of the conditional distribution of X given $C = c$, is well-defined unless c has zero probability. We set:

Definition 2. The *semiequivocation* of r.v. X given r.v. C is

$$S(X \mid C) = \sum \text{Prob}\{C = c\}\, S(X \mid C = c),$$

the sum being extended to all c's of positive probability.

Recall that the usual equivocation (conditional entropy) $H(X \mid C)$ can be defined in a similar way.

From the properties of the semientropies $S(X \mid C = c)$ one soon derives properties for the semiequivocation $S(X \mid C)$ (use corollary 1):

Corollary 2.

j) $S(X \mid C) \leq S(X)$; if X and C are independent $S(X \mid C) = S(X)$;

jj) $0 \leq S(X \mid C) \leq \log \frac{K}{2}$; $S(X \mid C) = 0$ iff for any cryptogram of positive probability there are at most two keys with positive conditional probability; for $K > 2$: $S(X \mid C) = log \frac{K}{2}$ iff for any such cryptogram the conditional probability of the random key is uniform.

The inequality in j), which is an essential requirement for any measure of conditional uncertainty, follows from concavity; note that the independence of X and C is not a necessary condition to have $S(X \mid C) = S(X)$: actually $S(X \mid C) = S(X)$ iff the conditional distributions of X given the cryptograms c of positive probabilities lie all on the same linearity segment (use theorem 2), or if they coincide, that is if X and C are independent. This is at variance with the case of the usual equivocation $H(X \mid C)$, where independence is also a necessary condition to have $H(X \mid C) = H(X)$. An explicit expression for $S(X \mid C)$ follows (use theorem 1).

Corollary 3. Set $h^*(p) = h(p)$ if $p \geq \frac{1}{2}$, $h^*(p) = \log 2$ else. Then

$$S(X \mid C) = H(X \mid C) + \sum \text{Prob}\{C = c\}\, h^*(\max_x \text{Prob}\{X = x \mid C = c\})$$

the sum being extended to all cryptograms c of positive probability and the max to all keys x.

We can now consider two functions of the non-negative integer n. Below C_n is the random cryptogram of length n made up of the first n random outputs of the cryptogram letter source. We use the *equivocation function* $e(n)$ and define a *semiequivocation function* $s(n)$:

$$e(n) = H(X \mid C_n), \qquad e(0) = H(X),$$

$$s(n) = S(X \mid C_n), \qquad s(0) = S(X).$$

It is known that $e(n)$ is non-increasing; using j) one obtains a similar property for $s(n)$. The corollary below lists also properties derived from corollary 1:

Corollary 4. The semiequivocation function $s(n)$ is a non-negative non-increasing function of n. One has:

$$e(n) - \log 2 \leq s(n) \leq e(n),$$

with equality on the left iff there are no keys with a conditional probability exceeding $\frac{1}{2}$, and equality on the right iff $e(n) = 0$.

Now we fix a "negligible" positive real number ε . We use the *unicity distance* d_1 and define a *duplicity distance* d_2. The former is the least integer for which $e(n) \leq \varepsilon$, the latter is the least integer for which $s(n) \leq \varepsilon$; if one or both of these integers do not exist, the corresponding distance is set equal to $+\infty$. As for their meaning, d_1 and d_2 represent the least number of cryptogram letters to be intercepted before the key equivocation, or the key semiequivocation, respectively, become negligible. If $d_1 = +\infty$, the cipher system with random key X and random cryptogram C_n is called (simply) *ideal*, if $d_2 = +\infty$ the cipher system is called *doubly ideal.* (Note that different definitions of unicity distance and ideal ciphers are found in the literature; the notions to be captured, however, are similar). As $s(n) \leq e(n)$, one has $d_2 \leq d_1$. In particular: any doubly ideal cipher is also simply ideal. The possibly void set of integers $\{n : s(n) \leq \varepsilon, e(n) > \varepsilon\} = \{n : d_2 \leq n < d_1\}$ is of relevance here: if the cryptogram length is in that set the cipher is unbreakable for a cryptanalyst who is devoid of "semantical competence" ($M = 1$), but is breakable for a cryptanalyst whose "semantical competence" is $M = 2$.

Example. Take a single-letter substitution cipher for a memoryless and stationary source (cf /1/ to /3/, /6/ or /8/). Assume that the cipher is complete (all $t!$ alphabet permutations are allowed to be used as keys, t being the number of distinct message letters in the message alphabet) and canonical (keys are equiprobable). Set:

$$A = t_1! t_2! \ldots t_r!$$

where r is the number of distinct components in the message letter p.d., each appearing $t_1, t_2, \ldots, t_r$ times, respectively ($t_1 + t_2 + \ldots + t_r = t$). One has $1 \leq A \leq t!$; $A = 1$ when all the t letter probabilities are distinct, $A = t!$ when the message letter p.d. is uniform. Then, for a suitable infinitesimal $\delta(n)$:

$$e(n) = H(X \mid C_n) = \log A + \delta(n)$$

(cf /1/ where more information on the asymptotic behaviour of $\delta(n)$ is given). This cipher has no asymptotic security for $A = 1$; in the sequel we assume that there are at least two source letters with the same probability. Then, for each key x and each cryptogram c, $\text{Prob}\{X = x \mid C_n = c\} = \text{Prob}\{X = \tilde{x} \mid C_n = c\} \leq \frac{1}{2}$, $\tilde{x}$ being the alphabet permutation obtained from x by interchanging those two letters. Therefore (corollary 4):

$$s(n) = e(n) - \log 2 = \log \tfrac{A}{2} + \delta(n)$$

In particular, for $A = 2$ (only two letters have the same probability) the cipher is simply ideal ($d_1 = +\infty$) and so cannot be broken however long the intercepted cryptogram is; instead, d_2 is finite (we assume $\varepsilon < \log 2$) and so, at least for sufficiently long cryptograms, the cipher can be broken by a semantically equipped cryptanalyst.

IV. A final comment

From the point of view of cryptographic applications our model based on the notions of semiequivocation and duplicity distance appears only as a mathematical abstraction: measuring the "semantical competence" of the cryptanalyst by an integer M, e.g. by $M = 2$, is certainly not a practical approach. On the other hand, in spite of all its drawbacks, the new model is more adequate than the classical one ($M = 1$), when the statistical description of the message source is not sufficiently robust so as to cover subtle and possibly long-range semantical dependences. The weakness of a frequency-type description has already been emphasized by exibiting certain paradoxes which it brings about (cf /6/ and /7/). Our new model serves as a warning against the dangers of using "clean" statistical message-source descriptions in cryptographic applications.

Appendix. We show that the system (2) has solutions when $p^* \leq \frac{1}{2}$. We proceed by induction on K. For $K = 2$ there is nothing to prove. For $K = 3$, $\aleph = \{a, b, c\}$, the system is solved by $R(a, b)$, $R(a, c)$ and $R(b, c)$ given by

$$R(x, y) = P(x) + P(y) - P(z),\ xyz = abc, acb, bca$$

Non-negativity holds since there is no single P-probability exceeding the sum of the other two ($p^* \leq \frac{1}{2}$; we have written $R(a, b)$ etc. instead of $R(\{a, b\})$ etc).

In the induction step from $K-1$ to K we shall blend the two smallest-probability letters, c and d, say; observe that, since $K \geq 4$, $P(c)+P(d)$ cannot exceed $\frac{1}{2}$. To improve readability we shall confine ourselves to describing the step from 3 to 4: it will be transparent that the restriction is only in the notation. We shall be contented with solutions with $R(c,d)=0$ and so the system to solve is:

$$\begin{aligned} R(a,b)+R(a,c)+R(a,d) &= 2P(a) \\ R(a,b)+R(b,c)+R(b,d) &= 2P(b) \\ R(a,c)+R(b,c) &= 2P(c) \\ R(a,d)+R(b,d) &= 2P(d) \end{aligned}$$

We blend c and d to form a super-letter $e=\{c,d\}$; we set $P(e)=P(c)+P(d)$, $R(x,e)=R(x,c)+R(x,d)$, $x=a,x=b$. The reduced system is as the one we have already solved for $M=3$, with e instead of c. We obtain a non-negative solution $R(a,b),R(a,e),R(b,e)$. Now we have to split $R(a,e)$ and $R(b,e)$ as the sum of two non-negative terms, $R(a,c)+R(a,d)$ and $R(b,c)+R(b,d)$ respectively, in such a way as to solve the unreduced system. As for the first two equations there (for the first $K-2$ equations in the generic induction step) any such non-negative splitting will do. As for the last two equations, a splitting as requested is feasible since we already know that one has

$$[R(a,c)+R(b,c)]+[R(a,d)+R(b,d)]=2[P(c)+P(d)]=2P(e)$$

QED

References

/1/ R. J. Blom, *Bounds on key equivocation for simple substitution ciphers*, IEEE Trans. Inform. Theory, vol. IT-25, pp.8-18, Jan. 1979

/2/ J. G. Dunham, *Bounds on message equivocation for simple substitution ciphers*, IEEE Trans. Inform. Theory, vol. IT-26, pp.522-527, Sept. 1980

/3/ A. Sgarro, *Error probabilities for simple substitution ciphers*, IEEE Trans. Inform. Theory, vol.IT-29, pp.190-198, March 1983

/4/ A. Sgarro, *Equivocations for homophonic ciphers*, in *Advances in Cryptology*, Proceedings of Eurocrypt 1984, pp. 51-61, Springer-Verlag, 1985

/5/ A. Sgarro, *Equivocations for transposition ciphers*, Rivista di matematica per le scienze economiche e sociali, Anno 8, fasc. 2, pp.107-114, 1985

/6/ A. Sgarro, *Exponential-type parameters and substitution ciphers*, Prbls. of Control and Inform. Theory, vol.14, pp. 393-403, 1985

/7/ A. Sgarro, *Information-theoretic versus decision-theoretic cryptography*, E und K, Sonderheft "Kryptologie und Datensicherheit", v.12, pp. 562-564, Springer–Verlag, 1987

/8/ H. Beker, F. Piper, *Cipher Systems*, Northwood Books, London, 1982

/9/ I. Csiszár, J. Körner, *Information Theory*, Academic Press, New York, 1982