# DIVERGENCE BOUNDS ON KEY EQUIVOCATION AND ERROR PROBABILITY IN CRYPTANALYSIS

Johan van Tilburg and Dick E. Boekee
Delft University of Technology
Department of Electrical Engineering
Information Theory Group
P.O. Box 5031, 2600 GA Delft, The Netherlands

## 0. Abstract

A general method, based on the f-divergence (Csiszar) is presented to obtain divergence bounds on error probability and key equivocation. The method presented here is applicable for discrete data as well as for continuous data. As a special case of the f-divergence it is shown that the upper bound on key equivocation derived by Blom is of the Bhattacharyya type. For a pure cipher model using a discrete memoryless message source a recursive formula is derived for the error probability. A generalization of the β-unicity distance is given, from which it is shown why the key equivocation is a poor measure of theoretical security in many cases, and why lower bounds on error probability must be considered instead of upper bounds. Finally the concept of unicity distance is generalized in terms of the error probability and is called the Pe-Security Distance.

## 1. Introduction

Cipher systems have given birth to the possibility of sending secret messages via public insecure channels. The secrecy of the messages depends highly on the strength of the cipher system used. When evaluating the theoretical strength of cipher systems, it is assumed that the cryptanalist behaves rationally, that he or she knows the set of transformations, the statistics of the message and the key source. The cryptanalist tries to estimate the message used and/or the key from the intercepted cryptogram. Shannon [1] used a probabilistic model for the theoretical analysis of secrecy systems. This model has been refined recently by Jürgensen and Matthews [2].

H.C. Williams (Ed.): Advances in Cryptology - CRYPTO '85, LNCS 218, pp. 489-513, 1986.

In Shannon's paper it is pointed out that if the cryptanalist intercepts a cryptogram, he is able to calculate the a posteriori probabilities of the various possible messages and keys which might have produced this cryptogram. This set of a posteriori probabilities describes how the cryptanalists knowledge of the message and key gradually becomes more precise as more enciphered text is intercepted. Shannon used as a measure of theoretical strength the equivocation which deals with a simplified description of the set of a posteriori probabilities. Then zero equivocation means that one key or message has a probability of one, and all others zero, corresponding to complete knowledge of the original key or message. Shannon also noticed that calculating the equivocation for the simplest type of cipher and language structure induces formulas which are nearly useless. His observation that the complexity of the problem suggests a method of approach, since sufficiently complicated problems can frequently be solved statistically, leads to the introduction of the famous "random cipher". Hellman [3] has shown that the random cipher actually defines a lower bound on the existence of good ciphers.

Blom [4] followed another way, by deriving an exponentially tight upper bound on the key equivocation for a simple substitution cipher (SSC) which is computationally more tractable. In Blom [5] an upper bound on the key equivocation for pure ciphers is given which exhibits the same structure as the bound in [4]. Later on, Dunham [6] derived bounds on the key appearance equivocation for an SSC and used the results of Blom [4] for bounding the message equivocation. Sgarro's paper [7] is based on an approach in coding theory, where one estimates error probabilities with respect to optimal coding problems. Sgarro made use of Kullback-Leibler divergence and composition classes to bound the error probability. His main results are asymptotic and contain the same relevant parameters as obtained by Blom [4] and Dunham [6].

## 2. Bounds with f-divergence

Typical in the approach is the use of information measures. For example, using Shannon's information measure leads to easy manipulation in a natural and intuitive way between different probability distributions (pd's). But still the underlying relevant parameter is the error probability (Pe). By bounding Pe with information measures, a region is determined in which the actual Pe can be found. The uncertainty in the value of Pe is resolved only in limiting cases where the bounds are tight. An excellent and straightforward use of this approach is given

by Lu [8], who uses Shannon's information measure to obtain the desired relation and applies the Fano inequality to lower bound Pe.

In the paper we consider the encipher system as a black box. Suppose we know the pd of the input (message) and the pd of the output (cryptogram). Transform the pd of the output to the input under a known key and compare the two pd's by means of Pe. Repeat this for all keys and select that key for which Pe is minimal. If ties occur then force a decision according to an arbitrary rule. Whereas determining Pe in a direct manner is quite involved, a much more natural way is to make use of the concept of distance measures since Pe is actually a distance measure itself.

The study of bounds on Pe has been of particular interest in the field of pattern recognition related to feature selection. Several distance measures have been used to obtain bounds on Pe, like Kolmogorov's variational distance, the Bhattacharyya distance, the J-divergence, the (generalized) Bayesian distance as well as many others. Much effort has been put into generalizing these measures from two classes with equal a priori probabilities to classes with non-equal a priori probabilities and from there to m classes, with $m \geq 2$. The comparison of the various bounds on Pe has also received much attention. More details can be found in Kanal [9] and in Chen [10].

A generalized approach can be given by using the f-divergence, as defined by Csiszar [11]. In this paper we shall use a slight modification, which we shall call the normalized average f-divergence. This divergence measure is directly related to Pe by its very definition, and it is therefore convenient for manipulating in this theoretical context. We shall use a definition which is sufficient for this paper. More details can be found in Boekee and van Tilburg [12] and in [13][14].

Before continuing, a short note about the notation. As far as possible the notation is in agreement with that of Blom in [4][5], with the exception that the logarithms involved are taken to base 2. Throughout this paper we shall use the convention that capital letters denote random variables, boldface letters denote sequences, capital script letters are reserved for sets and lower case letters represent the elements in a set.
Let $\mathcal{S}$ denote an arbitrary (finite) set with cardinal number $|\mathcal{S}|$. $\mathcal{S}^L$ is the class of all sequences $\mathbf{s}$ of length L. A sequence (concatenation of symbols) $\mathbf{s}$ of length L of elements s (not necessarily different) in $\mathcal{S}$ is indicated by $\mathbf{s}^L$.
The cipher model is a set of uniquely reversible transformations of

$T = \{t_j\}_{j=1}^{J}$ of a set of possible messages $M = \{m_n\}_{n=1}^{N}$ into a set of cryptograms $E = \{e_n\}_{n=1}^{N}$, the transformations having associated probabilities $P = \{p_j\}_{j=1}^{J}$. $J$ is the cardinal number of the set of keys $K = \{k_j\}_{j=1}^{J}$.

Definition 2.1: The normalized average f-divergence (for short: $f$-divergence) for isonorm-functions $f(x)$ is given by:

$$\bar{D}_f = \bar{D}_f(1;2) = \frac{f_\infty - \bar{D}_f(1;2)}{2.f_\infty - f_1},$$

where f is a convex function satisfying:

$$f_\infty = \lim_{x\to\infty} \frac{f(x)}{x}, \qquad f_0 = \lim_{x\downarrow 0} f(x),$$

$$f_0 = f_\infty \quad \text{(isonorm restriction)},$$

$$f_1 = f(1)$$

and

$$\bar{D}_f(1;2) = E_{E^L}\left[ f\left(\frac{P_{K/E^L}(k_1/e^L)}{P_{K/E^L}(k_2/e^L)}\right) \cdot P_{K/E^L}(k_2/e^L) \right]$$

is the average f-divergence for discrimination of key $k_1$ against $k_2$. By $E_{E^L}$ we mean the expectation operator. □

If we define $f_*(x) = x.f(\frac{1-x}{x})$ and $u^L = u(e^L) = P_{K/E^L}(k_2/e^L)$, then it follows that:

$$f(u^L) = \tfrac{1}{2}\,\frac{f_*(0) - f_*(u^L)}{f_*(0) - f_*(\tfrac{1}{2})},$$

and hence $\bar{D}_f = E_{E^L}[f(u^L)]$.

Note that $Pe = Pe(K/E^L) = 1 - E_{E^L}[\max(u^L, 1-u^L)] = E_{E^L}[\min(u^L, 1-u^L)]$, which shows that the $f$-divergence includes the error probability as a special case for $f(x) = Pe(x) = \min(x, 1-x)$.

Definition 2.2: The Bhattacharyya distance is given by

$$B = -\log \rho,$$

where

$$\rho = \rho(E^L/K) = \sum_{e^L \in E^L} \sqrt{p_{E^L/K}(e^L/k_1).p_{E^L/K}(e^L/k_2)}$$

is the Bhattacharyya coefficient. □

If we take the a priori probabilities of the keys into account, we obtain the next definition.

Definition 2.3. The average Bhattacharyya distance is given by

$$\bar{B} = - \log \bar{\rho},$$

where

$$\bar{\rho} = \bar{\rho}(K/E^L) = E_{E^L}[\sqrt{P_{K/E^L}(k_1/e^L) \cdot P_{K/E^L}(k_2/e^L)}]$$

$$= \sum_{e^L \in E^L} \sqrt{P_{KE^L}(k_1, e^L) \cdot P_{KE^L}(k_2, e^L)}$$

is the average Bhattacharyya coefficient. □

If the keys are equiprobable it follows that $\bar{\rho}(K/E^L) = \frac{1}{2}\, \rho(E^L/K)$. If we set $f(x) = -x^{1-a}$, we find that $f(x) = x^a \cdot (1-x)^{1-a}$. Then the $f$-divergence becomes

$$\bar{D}_f = E_{E^L}[f(u^L)] = E_{E^L}[(u^L)^a \cdot (1-u^L)^{1-a}],$$

which is the Chernoff distance $C_a(K/E^L)$. For $a=\frac{1}{2}$ we have

$$\bar{D}_f = E_{E^L}[\sqrt{u^L \cdot (1-u^L)}] = E_{E^L}[\sqrt{P_{K/E^L}(k_1/e^L) \cdot P_{K/E^L}(k_2/e^L)}],$$

which shows that the average Bhattacharyya coefficient is a special case of this $f$-divergence.
Similarly we find for $f(x) = |1-x^{1/r}|^r$:

$$\bar{D}_f = \tfrac{1}{2} - \tfrac{1}{2} \cdot \bar{M}_r^{\,r},$$

where

$$\bar{M}_r = [E_{E^L}\{|P_{K/E^L}(k_1/e^L)^{1/r} - P_{K/E^L}(k_2/e^L)^{1/r}|^r\}]^{1/r}$$

is the generalized Matusita distance.
For r=1 we have Kolmogorov's variational distance and for r=2 the usual Matusita distance.
In the next theorem a class of upper and lower bounds on the $f$-divergence is considered in terms of Pe. A sufficient condition for the validity of the theorem is to restrict the $f$-divergence to symmetric functions, i.e. $f(1-u)=f(u)$.

Theorem 2.1. A class of upper and lower bounds induced by Pe on the symmetric $f$-divergence is

$$Pe \leq \bar{D}_f \leq f(Pe).$$

Proof: First observe that $f(x)$ is a normalized concave function on $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$ resp., such that $f(x) \geq \min(x,1-x)$ for $x \in [0,1]$ with equality at least for $x \in \{0,\frac{1}{2},1\}$:

i) since $\min(u^L,1-u^L) \leq f(u^L)$ it follows that

$$\bar{D}_f = E_{E^L}[f(u^L)] \geq E_{E^L}[\min(u^L,1-u^L)] = Pe,$$

ii) $$\bar{D}_f = E_{E^L}[f(u^L)] = E_{E^L}[f(\min(u^L,1-u^L))] \leq f(E_{E^L}[\min(u^L,1-u^L)]) = f(Pe).$$ □

Remark. The theorem also gives bounds for normalized concave functions $f(x)$which do not satisfy the $f$-divergence. Moreover the symmetric restriction is not used in the proof of the lower bound. So for the Chernoff bound it holds that $Pe \leq C_a(K/E^L)$.

In fact, the lower bound stated is a direct upper bound on $Pe(K/E^L)$. The upper bound in this theorem sometimes cannot be rewritten (explicitly)as a bound on $Pe(K/E^L)$. This can be a disadvantage if we are interested in bounds on $Pe(K/E^L)$. However, the lower bound on $Pe(K/E^L)$ can then be computed numerically or indirectly via the upper bound.

Example 2.1. Bounds for the average Bhattacharyya coefficient. Because then $f(x) = \sqrt{x.(1-x)}$, we obtain

$$Pe \leq \bar{\rho}(K/E^L) \leq \sqrt{Pe.(1-Pe)}$$

and

$$\tfrac{1}{2}.(1 - \sqrt{(1 - 4\,\bar{\rho}^2)}\,) \leq Pe(K/E^L) \leq \bar{\rho}.$$ □

Example 2.2. Bounds for the key equivocation. Because then $f(x)=\frac{1}{2}.h(x)=\frac{1}{2}.[-x.\log x - (1-x).\log(1-x)]$, we obtain

$$Pe \leq \tfrac{1}{2}.H(K/E^L) \leq \tfrac{1}{2}.h(Pe),$$

where

$$H(K/E^L) = E_{E^L}[h(u^L)]$$

is Shannon's key equivocation.

□

The next lemma can easily be verified. A proof can be found, e.g., in Ito [15]. The lemma is illustrated in figure 2.1.

Lemma 2.1. The relation between the bounds is given by

$$Pe(K/E^L) \leq \tfrac{1}{2}.H(K/E^L) \leq \bar{p}(K/E^L).$$

□

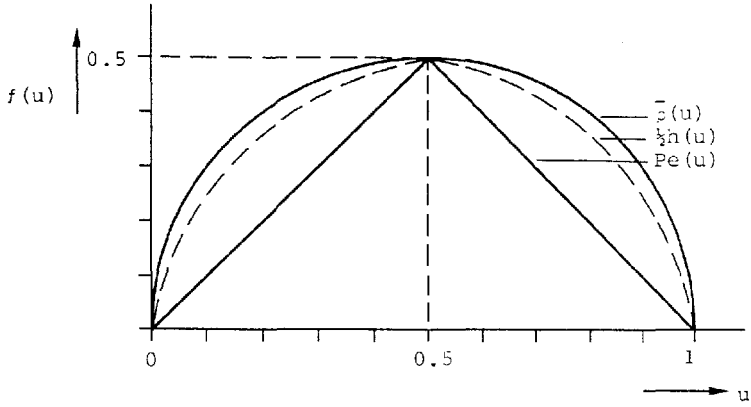


Figure 2.1. The basic functions which constitute the measures $\bar{p}$, $\frac{1}{2}H$ and Pe.

Next $\bar{p}$ will be determined for the general case, after which we shall return to the binary case. The model used is that of a pure cipher with the following assumptions.

- The message and the key are stochastically independent.
- The message source is discrete and memoryless.

$$\begin{pmatrix} M \\ Q_M \end{pmatrix} = \begin{pmatrix} 1 & 2 & \cdots\cdots & N \\ q_1 & q_2 & \cdots\cdots & q_N \end{pmatrix}.$$

- $T$ is the set of all unique invertible transformations $t_j$ of $M$ onto $E$, where the index $j$ is the associated key.

$$T = \{t_j\}_{j=1}^{J}.$$

  Note that $|K| = J$.
- The cryptogram alphabet $E$ is (not necessarily) identical to $M$.

$$\begin{pmatrix} E \\ Q_{t_j^{-1}(E)} \end{pmatrix} = \begin{pmatrix} 1 & 2 & \cdots\cdots & N \\ q_{t_j^{-1}(1)} & q_{t_j^{-1}(2)} & \cdots\cdots & q_{t_j^{-1}(N)} \end{pmatrix}.$$

Definition 2.4. Pure cipher (see Blom [16, theorem 3]).
A cipher is pure if and only if its set of enciphering transformations $T$ is a coset (left or right) in $G$ and

the keys are equiprobable. □

Remark. $G$ is the multiplicative group of all invertible transformations of $M$ onto $M$.

Following Blom [5], the set of enciphering transformations $T$ forms a left coset in the group $G$. When the keys are equiprobable, it follows that the cipher is pure.

As $T$ is a left coset, we may define

$$T = \{gr_j\}_{j=1}^{J} ,$$

where $g \in G$ and $R = \{r_j\}_{j=1}^{J}$ is a subgroup in $G$.

Recalling Shannon [1], two secrecy systems R and S are similar, if there exists an invertible transformation A such that R = A.S. This means that enciphering with R is the same as enciphering with S followed by the transformation A. It is clear that similarity is an equivalence relation. The problem of finding bounds for a cipher using the set of transformations $T$ is now transformed to a cipher using the set of transformations $R$, where $R$ is a subgroup in $G$.

Theorem 2.2. The Bhattacharyya coefficient for the $j^{th}$ transformation in a pure cipher model $T$ using an N-ary discrete memoryless source with a priori probabilities $q_n$ is given by

$$\rho_{1j} = \left( \sum_{n=1}^{N} \sqrt{q_n \cdot q_{r_j^{-1}(n)}} \right)^{L},$$

where

$r_j^{-1} \in R$, $R$ is the group generating $T$.

$r_1$ is the identity element in $R$.

Proof. Because the keys are equiprobable and independent of the message source it follows for L=1 that

$$p_{EK}(e,k_i) = \frac{1}{J} \cdot p_M(r_{k_i}^{-1}(e)),$$

where

$$p_M(x) = \sum_{n=1}^{N} q_n \cdot \delta(n-x).$$

As stated we compare the pd of the message source with the inversely transformed encryption pd which depends on the transformation (key) used. By noting that $k_1$ is associated with the identity transformation, $\bar{\rho}$ becomes

$$\bar{\rho}_{1j} = \sum_{e \in E} \sqrt{p_{EK}(e,k_1) \cdot p_{EK}(e,k_j)} \ .$$

after substitution of $p_{EK}(e,k_i)$ one easily obtains

$$\bar{\rho}_{1j} = \frac{1}{J} \sum_{n=1}^{N} \sqrt{q_n \cdot q_{r_j^{-1}(n)}} \ .$$

$\bar{\rho}_{1j}$ for the extension of the cryptogram (L > 1) follows directly from the weak additivity of the Bhattacharyya distance B.

$$- L.\log \rho = - \log \rho^L,$$

so that

$$\rho_{1j} = J \cdot \bar{\rho}_{1j} = \Big( \sum_{n=1}^{N} \sqrt{q_n \cdot q_{r_j^{-1}(n)}} \Big)^L. \qquad \square$$

For the binary case $\bar{\rho}_{1j}$ reduces to $\bar{\rho}_{12} = \bar{\rho} = \frac{1}{2}(\sqrt{4q_1q_2})^L$.

Substitution in lemma 2.1 and example 2.1 proves the next theorem.

Theorem 2.3. Bounds on the average probability of error (or probability of incorrect key identification) in a pure cipher model using a discrete memoryless source with a priori probabilities $q_n$ are:

$$\tfrac{1}{2}.(1-\sqrt{(1-(4q_1q_2)^L)}) \leq Pe(K/E^L) \leq \tfrac{1}{2}.H(K/E^L) \leq \tfrac{1}{2}.(\sqrt{4q_1q_2})^L. \qquad \square$$

The upper bound on the key equivocation is the same as obtained by Blom [4] using an SSC-model; however, at the same time we have a lower bound too. Moreover, for a different cipher model we only have to substitute the corresponding $\bar{\rho}$ in example 2.1, yielding the new upper and lower bounds. This illustrates the general structure of the bounds.

By a similar argument it can be shown that for the Chernoff bound it holds that

$$Pe(K/E^L) \leq C_a(K/E^L) = \tfrac{1}{2}.(q_1^a.q_2^{1-a} + q_2^a.q_1^{1-a})^L,$$

where $0 \leq a \leq 1$.
This is a symmetric upper bound, which is minimal for $a=\frac{1}{2}$; that is, if it coincides with the Bhattacharyya bound. This shows that the Bhattacharyya bound is optimal in this context.

Thus far bounds on Pe have been considered. In the next theorem some recursive properties of Pe are stated.

Theorem 2.4. For the average probability of error (or probability of incorrect key identification) in a pure cipher model using a discrete memoryless source with a priori probabilities $p \geq q$ it holds that

i) L is even: $Pe(K/E^{L+1}) = Pe(K/E^{L}) - \frac{1}{2}(p-q)\binom{L}{L/2}(\sqrt{pq})^{L}$

with $Pe(K/E^{0}) = 0.5$

ii) L is odd : $Pe(K/E^{L+1}) = Pe(K/E^{L})$.

Proof

i) If L is even we have

$$Pe(K/E^{L+1}) = \sum_{i=0}^{L+1}\left[\binom{L+1}{i}p^{i}q^{L+1-i}.\min\left(\frac{p^{i}q^{L+1-i}}{p^{i}q^{L+1-i}+p^{L+1-i}q^{i}}, \frac{p^{L+1-i}q^{i}}{p^{i}q^{L+1-i}+p^{L+1-i}q^{i}}\right)\right]$$

$$= \sum_{i=0}^{L/2}\left[\binom{L+1}{i}p^{i}q^{L+1-i}\right] = q \cdot \sum_{i=0}^{L/2}\left\{\left[\binom{L}{i} + \binom{L}{i-1}\right] \cdot p^{i}q^{L-i}\right\}$$

$$= q.Pe(K/E^{L}) + \frac{1}{2}\binom{L}{L/2}(pq)^{L/2} + p \cdot \sum_{i=0}^{L/2-1}\left[\binom{L}{i}p^{i}q^{L-i}\right]$$

$$= Pe(K/E^{L}) - \frac{1}{2}(p-q)\binom{L}{L/2}(pq)^{L/2}.$$

ii) if L is odd we have

$$Pe(K/E^{L+1}) = \sum_{i=0}^{\frac{L+1}{2}}\left[\binom{L+1}{i}p^{i}q^{L+1-i}\right] - \frac{1}{2}\binom{L+1}{\frac{L+1}{2}}(pq)^{L/2}$$

$$= q \cdot \sum_{i=0}^{\frac{L+1}{2}}\left[\binom{L}{i}p^{i}q^{L-i}\right] + q \cdot \sum_{i=1}^{\frac{L+1}{2}}\left[\binom{L}{i-1}p^{i}q^{L-i}\right] - \frac{1}{2}\binom{L+1}{\frac{L+1}{2}}(pq)^{L/2}$$

$$= q.Pe(K/E^{L}) + \binom{L}{\frac{L+1}{2}}(pq)^{\frac{L+1}{2}} + p \cdot \sum_{i=1}^{\frac{L-1}{2}}\left[\binom{L}{i}p^{i}q^{L-i}\right] - \frac{1}{2}\binom{L+1}{\frac{L+1}{2}}(pq)^{L/2}.$$

Since $\frac{1}{2}\cdot\binom{L+1}{\frac{L+1}{2}} = \binom{L}{\frac{L+1}{2}}$ it follows that $Pe(K/E^{L+1}) = Pe(K/E^{L})$. □

An efficient algorithm can be obtained if theorem 2.4 is written in the following way.

$Pe(0) = 0.5,$
$A(0) = p - 0.5,$
$B = 4.p.(1-p),$

for L is even: $Pe(L+2) = Pe(L+1) = Pe(L) - A(L)$
$A(L+2) = A(L).(1-1/(L+2)).B.$

Remark. Although the key equivocation is a simplified description of the set of a posteriori probabilities, it cannot be transformed into an effective algorithm of the above type for an SSC-model.

From the theorem we may conclude that the behaviour of Pe for small L is determined by $|p-q|$. This is in contrast to the longterm exponential behaviour which is characterized by $|\sqrt{p} - \sqrt{q}|$.

A discussion of the relations between the different bounds and the actual average error probability is deferred to section 4, where the variance of the bound is investigated too.

## 3. Bound extensions of the Bhattacharyya type

In section 2 we have considered bounds on Pe for cipher systems using binary sources. We now turn to the N-ary problem where a general bound in terms of the $f$-divergence can be given. Generally speaking this generalized bound becomes less tight for increasing N. However, some particular functions allow better bounds. For this reason this section is completely devoted to the extension of the Bhattacharyya bound. Furthermore there exists a general class of distance measures (for instance the general mean distance [14]) which are inherently based on the N-ary problem. First, a general bound of the Bhattacharyya type is derived on $Pe(K/E^L)$ as well as on $H(K/E^L)$, after which the bound is restricted to the pure cipher model. Finally, the pure cipher bound is applied in the case of a discrete memoryless source.

A direct extension of the Bhattacharyya bound can be found by making use of the following theorem:

Theorem 3.1. The upper bounds on the probability of error in a cipher model using an N-ary source are given by

$$Pe(K/E^L) \leq \tfrac{1}{2}.H(K/E^L) \leq \tfrac{1}{2}.\log \left( \sum_{i=1}^{J} \sum_{j=1}^{J} \bar{\delta}_{ij} \right),$$

where

$$\bar{\delta}_{ij} = E_{E^L} \left[ \sqrt{P_{K/E^L}(k_i/e^L).P_{K/E^L}(k_j/e^L)} \right].$$

Proof. i) Kovalevski [17] has proved that

$\log m + m.(m+1).\log(\frac{m+1}{m}).(Pe(K/E^L) - \frac{m-1}{m}) \leq H(K/E^L)$, with $\frac{m-1}{m} \leq Pe \leq \frac{m}{m+1}$.

For m=1 or $0 \leq Pe \leq \frac{1}{2}$ this bound reduces to $2.Pe(K/E^L) \leq H(K/E^L)$. This holds for $Pe > \frac{1}{2}$ too, with the implication that the implicit expression is tighter.

ii) $$H(K/E^L) = E_{E^L}\left[H(K/e^L)\right] = E_{E^L}\left[-\sum_{i=1}^{J} P_{K/E^L}(k_i/e^L).\log P_{K/E^L}(k_i/e^L)\right]$$

$$\leq E_{E^L}\left[\log \left(\sum_{i=1}^{J} P_{K/E^L}(k_i/e^L)\right)^2\right]$$

$$= E_{E^L}\left[\log \sum_{i=1}^{J}\sum_{j=1}^{J} \sqrt{P_{K/E^L}(k_i/e^L).P_{K/E^L}(k_j/e^L)}\right]$$

$$\leq \log \sum_{i=1}^{J}\sum_{j=1}^{J} \bar{\rho}_{ij}.$$

This result has implicitly been proved by Blom [4].

Combining (i) and (ii) yields the theorem. □

For the pure cipher model we have

. independence of keys used: $\sum_{j=1}^{J} \bar{\rho}_{ij} = \sum_{j=1}^{J} \bar{\rho}_{kj}$ ,

. equiprobable keys : $\bar{\rho}_{ij} = \frac{1}{J}$ for i=j,

which implies the next corollary.

Corollary 3.1. For the upper bounds on the probability of error in a pure cipher model using an N-ary source, we have

$$Pe(K/E^L) \leq \tfrac{1}{2}H(K/E^L) \leq \tfrac{1}{2}\log(1 + \sum_{i=2}^{J} \rho_{1j}),$$

where

$$\rho_{1j} = J.\bar{\rho}_{1j} \text{ and } J = |K|.$$ ■

The next corollary ensues from substituting $\rho_{1j}$ (Theorem 2.2) in corollary 3.1. Because the summation of $\rho_{1j}$ is taken over all transformations in the group R it makes no difference if we write $r_j(n)$ instead of $r_j^{-1}(n)$.

orollary 3.2. For the upper bounds on the probability of error in a pure cipher model $T$ usingan N-ary discrete memoryless source, we have

$$\text{'Pe}(K/E^L) \leq \tfrac{1}{2}H(K/E^L) \leq \tfrac{1}{2}.\log\left[1 + \sum_{j=1}^{J}\left(\sum_{n=1}^{N}\sqrt{q_n \cdot q_{r_j(n)}}\right)^L\right],$$

here

$j \in R$, $R$ is the group generating $T$ and $r_1$ is the identity element. □

he upper bound on the key equivocation is the same as obtained by Blom 5]. However, the proof is simplified considerably and the general struc-ure of the bound becomes clear.

. lower bound can be found by using the natural multiplicative exten-ion of the Bhattacharyya coefficient. A general (non-trivial) upper ound for this extension does not exist (van Tilburg [18]).

heorem 3.2. A lower bound on the probability of error in a cipher mo-del using an N-ary source is given by

) $[1 - Pe(K/E^L)] \cdot Pe(K/E^L)^{J-1} \geq (J-1)^{J-1} \cdot \bar{\rho}_J^{\,J}$,

i) $Pe(K/E^L) \geq (J-1) \cdot \bar{\rho}_J^{\frac{J}{J-1}}$,

here $\bar{\rho}_J = E_{E^L}\left[\prod_{j=1}^{J} P_{K/E^L}(k_j/e^L)^{1/J}\right]$ is the multiplicative extension of he average Bhattacharyya coefficient.

roof. i) Define $x = x(e^L) = \max_j\left[P_{K/E^L}(k_j/e^L)\right]$. Then

$$\bar{\rho}_J = E_{E^L}\left[\prod_{j=1}^{J} P_{K/E^L}(k_j/e^L)^{1/J}\right] \leq E_{E^L}\left[x^{1/J} \cdot \left(\frac{1-x}{J-1}\right)^{\frac{J-1}{J}}\right]$$

$$\leq \left(E_{E^L}[x]\right)^{1/J} \cdot \left(E_{E^L}\left[\frac{1-x}{J-1}\right]\right)^{\frac{J-1}{J}} = \left(1-Pe(K/E^L)\right)^{1/J} \cdot \left(\frac{Pe(K/E^L)}{J-1}\right)^{\frac{J-1}{J}}$$

$$\text{or } (J-1)^{J-1} \cdot \bar{\rho}_J^{\,J} \leq (1-Pe(K/E^L)) \cdot Pe(K/E^L)^{J-1}.$$

.i) Simplifying the inequality in (i) by making use of

$$(1-Pe) \cdot Pe^{J-1} \leq Pe^{J-1}$$

.mplies

$$(J-1) \cdot \bar{\rho}_J^{\left(\frac{J}{J-1}\right)} \leq Pe(K/E^L).$$ □

The proof of the next lemma is similar to that of theorem 2.2. and is therefore omitted.

Lemma 3.1. The multiplicative extension of the Bhattacharyya coefficient in a pure cipher model $T$ using an N-ary discrete memoryless source with a priori probabilities $q_n$ is given by

$$\rho_J = \left[ \sum_{n=1}^{N} \prod_{j=1}^{J} q_{r_j(n)}^{1/J} \right]^L,$$

where

$r_j \in R$, $R$ is the group generating $T$ and $\rho_J = J \cdot \bar{\rho}_J$. □

Substituting lemma 3.1 in theorem 3.1 yields corollary 3.3.

Corollary 3.3. A lower bound on the probability of error in a pure cipher model $T$ using an N-ary discrete memoryless source with a priori probabilities $q_n$ is given by

$$\frac{J-1}{J^{\left(\frac{J}{J-1}\right)}} \cdot \left[ \sum_{n=1}^{N} \prod_{j=1}^{J} q_{r_j(n)}^{1/J} \right]^{\left(\frac{J}{J-1}\right).L} \leq Pe(K/E^L),$$

where

$r_j \in R$ and $R$ is the group generating $T$. □

For large-sized key spaces we have the tight approximation

$$\left[ \sum_{n=1}^{N} \prod_{j=1}^{J} q_{r_j(n)}^{1/J} \right]^L \leq Pe(K/E^L).$$

## 4. The Pe-security distance

The $f$-divergence is defined in a probabilistic environment and therefore easily fits into the probabilistic model of cryptosystems proposed by Jürgensen and Matthews [2]. In their paper (section 6) they have defined the $\beta$-UD as $\underset{L}{MIN} \{L \mid H(K/E^L) \leq \beta\}$.
They also propose the $(\alpha,\beta)$-security distance: a system is said to be $(\alpha,\beta)$-secure at L if $Pr\{H(K/e^L) \leq \beta\} \leq \alpha$. In the present section the $\beta$-UD is related to Pe and is not restricted to the key equivocation only. To avoid confusion we refer to this generalized $\beta$-UD as the $\gamma$-UD. When discussing the results of the SSC-model it is observed that Pe is a natural (theoretical) security measure. By noting this, Pe is derived

for a random cipher (RC-model). As a result it is found that at unicity distance Pe highly depends on the size of the key space. Hence we conclude that linking the UD to Pe leads to a better and more adequate explanation of the unicity distance. Finally, the concept of UD is generalized in terms of Pe and is called the Pe-security distance (Pe-SD). This security distance can be considered as a special case of the $\gamma$-UD and includes the original UD found in an RC-model too. Moreover, it becomes clear that lower bounds are needed to approximate the Pe-SD. For the key equivocation this means that one must make use of the Fano-inequality, because the key equivocation itself defines an upper bound.

In this section our main concern is the binary case. For this reason and to avoid unnecessary notational problems the $\gamma$-UD is mainly described for the binary case.

Definition 4.1. The generalized $\gamma$-unicity distance or, for short, the $\gamma$-UD is defined as

$$L(\gamma) = \underset{L}{\text{MIN}}\{L \in \mathbb{R}^+ \mid E_{E^L}[g(u^L)] \leq \gamma\},$$

where

$$u^L = u(e^L) = P_{K/E^L}(k_2/e^L),$$

and $g(.)$ is a normalized function such that $g(x) \geq \min(x,1-x)$ for $x \in [0,1]$ with equality at least for $x \in \{0,\frac{1}{2},1\}$.

□

Observe that for $g(u^L) = f(u^L)$ we have the $\gamma$-UD for the $f$-divergence, whereas if $g(u^L) = \frac{1}{2}.h(u^L)$ the $\gamma$-UD for the normalized key equivocation is obtained.

The $\gamma$-UD not only depends on the measure used, but it depends on the model used (including the source) too. This is illustrated by the following examples:

Example 4.1. For the $\gamma$-UD using the key equivocation we have

$$L(\gamma) = \underset{L}{\text{MIN}}\{L \in \mathbb{R}^+ \mid \tfrac{1}{2}H(K/E^L) \leq \gamma\}.$$

If the key and message sources are independent we find

$$L(\gamma) = \underset{L}{\text{MIN}}\{L \in \mathbb{R}^+ \mid H(E^L) - H(\mathbf{M}^L) \geq H(K) - 2.\gamma\}.$$

If in addition the message source is memoryless this becomes

$$L(\gamma) = \underset{L}{\text{MIN}}\{L \in \mathbb{R}^+ \mid L \geq L(0) \cdot (1 - \frac{2\gamma}{H(K)})\},$$

with

$$L(0) = \frac{H(K)}{H(E) - H(M)} .$$

Note that $H(K/\mathbf{E}^L)$ is convex in the sense that $H(K/\mathbf{E}^L) - H(K/\mathbf{E}^{L+1}) \geq H(K/\mathbf{E}^{L+1}) - H(K/\mathbf{E}^{L+2})$ and $H(E) - H(M) = H(K/E^0) - H(K/E^1)$, so that $L(0)$ can be found as the point of intersection of the straight line through $H(K/E^0)$ and $H(K/E^1)$ with the L-axis. This line defines a lower bound on the key equivocation. □

Example 4.2. For a random cipher model we have

$$L(\gamma) = \underset{L}{\text{MIN}}\{L \in \mathbb{R}^+ \mid L \geq L(0) \cdot (1-2\gamma)\},$$

with

$$L(0) = \frac{\log |K|}{\log |M| - H_L(M)} .$$

$L(0)$ is the original UD for the RC-model obtained by Shannon [1]. $H_L(M)$ denotes the entropy per symbol in a sequence of L message symbols, i.e. $H_L(M) = H(\mathbf{M}^L)/L$. Note that the decrease of $L(\gamma)$ is linear; this is not necessarily so for other models. □

Now we are able to state the next lemma which is a generalization of the second part of proposition 7.6 [2] with the assumptions made above. The proof is similar and is therefore omitted.

Lemma 4.1. If $L_0$ is the $\gamma$-unicity distance, then for $L > L_0$ we have

$$\Pr\{g(u^L) \leq \gamma\} \geq \frac{\underset{\mathbf{E}^L}{\text{Var}}[g(u^L)]}{\gamma^2 + \underset{\mathbf{E}^L}{\text{Var}}[g(u^L)]} .$$

■

Let us consider an SSC-model using a binary memoryless message source with a priori probabilities p=0.6 and q=0.4. The upper and lower bounds on Pe (derived in section 2) are applied to this model and illustrated in figure 4.1, in which the exact value of Pe is given too. The figure shows that $\bar{\rho}$ and $\frac{1}{2}H$ are loose upper bounds for small values of L; even at UD they are still not tight. This is also demonstrated by the next example.

Example 4.3. (see also example 7.2 [2]).

Consider the SSC-model with p=0.7 and L=7 (RC-model: UD=8.4). For this model we have $\bar{\rho}(K/E^7) = 0.27$ and $\frac{1}{2}.H(K/E^7) = 0.22$.
Although these values are not too high they are still much too optimistic since $Pe(K/E^7) = 0.126$. □

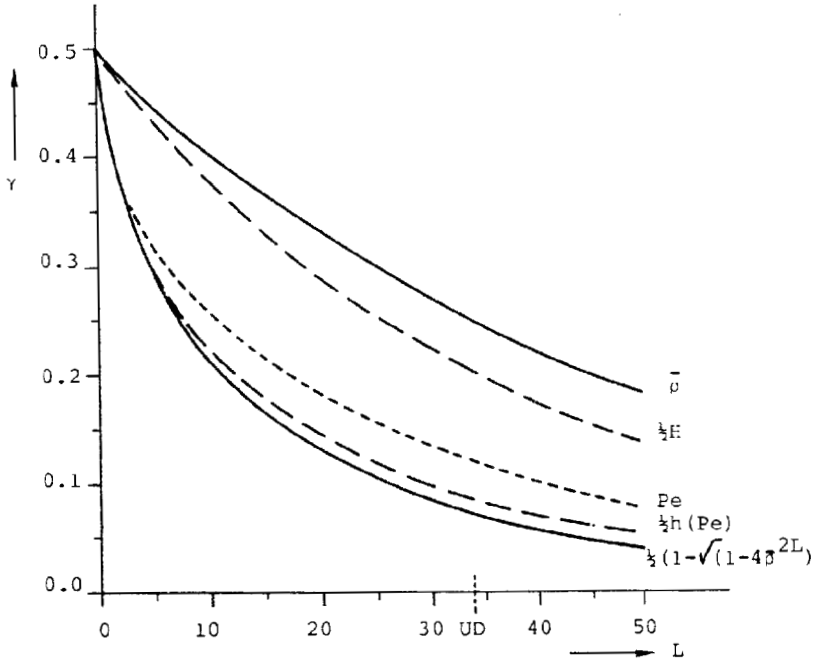


Figure 4.1. Bounds on the average probability of incorrect key identification Pe in a memoryless SSC-model with p=0.6.

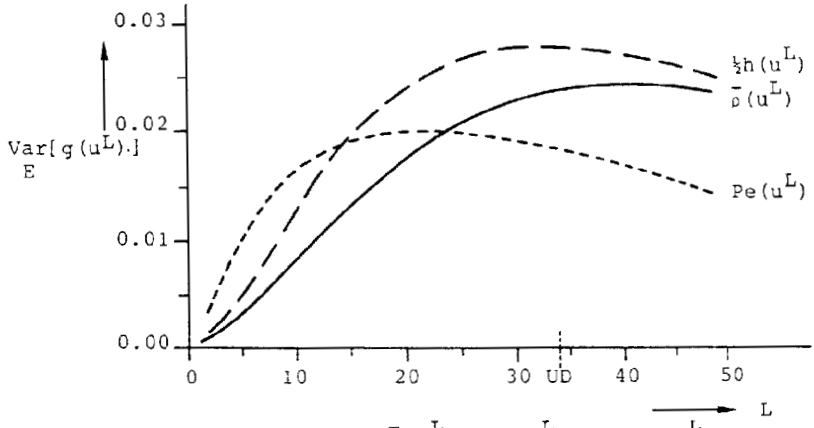


Figure 4.2. The variance of $\bar{\rho}(u^L)$, $\frac{1}{2}h(u^L)$ and $Pe(u^L)$ for a memoryless SSC-model with p=0.6.

In addition to all this, consider $L(\gamma)$ with $\gamma$ constant for the different bounds. Now it becomes clear that the $\beta$-UD (and thus the $(\alpha,\beta)$-SD, too) is a poor and positively biased estimator of L in $Pe(K/E^L)=\gamma$; this in contrast to the lower bounds which are negatively biased and tighter.

In figure 4.2. the variance of $g(u^L)$ is shown for the SSC-model. It is observed that the variance of $H(K/e^L)$ is maximal at UD, which is found for other values of p too. Moreover, the length of the cipher text at which the variance of $Pe(u^L)$ reaches its maximum is always less than the length obtained by $\bar{\rho}(u^L)$ and $h(u^L)$. This can be explained from the convex nature of $Pe^2(u^L)$; this in contrast to $\bar{\rho}^2(u^L)$, which is a concave function (see also figure 4.3). Besides this, for the normalized functions it holds that $Var[g(u^L)] \leq E[g(u^L)]$ since $Var[g(u^L)] \leq E[g^2(u^L)] \leq E[g(u^L)]$. This is illustrated by the next example.

Example 4.4. (see also p. 292 [3] and p. 343 [2]).
Suppose that after intercepting L enciphered symbols it holds that

$$n_k \begin{cases} 0 & \text{with probability } 1 - 10^{-10} \\ 10^{20} & \text{with probability } 10^{-10}, \end{cases}$$

in which $n_k$ is the number of spurious key decipherments. Then $\bar{n}_k = 10^{10}$ and $Var(n_k) \approx 10^{30}$.

In the worst case $P_{K/E^L}(k_i/e^L) = P_{K/E^L}(k_j/e^L)$ for all $k_i$ and $k_j$ in K, so that the key space must satisfy $|K| = 10^{20}+1$. For Pe we then obtain

$$Pe(K/E^L) \approx 10^{-10} \quad \text{and} \quad Var_{E^L}[Pe(u^L)] \approx 10^{-10}.$$

Since the real key space may be larger, say for example $|K| = 10^{100}$ it follows that

$$Pe(K/E^L) \approx 10^{-90} \quad \text{and} \quad Var_{E^L}[Pe(u^L)] \approx 10^{-170}.$$

So, the interpretation of $\bar{n}_k$ (and $H(K/E^L)$ also) depends greatly on the size of the key space. For this reason it is necessary to utilize normalized functions. Moreover, the interpretation of the variance becomes more realistic too. □

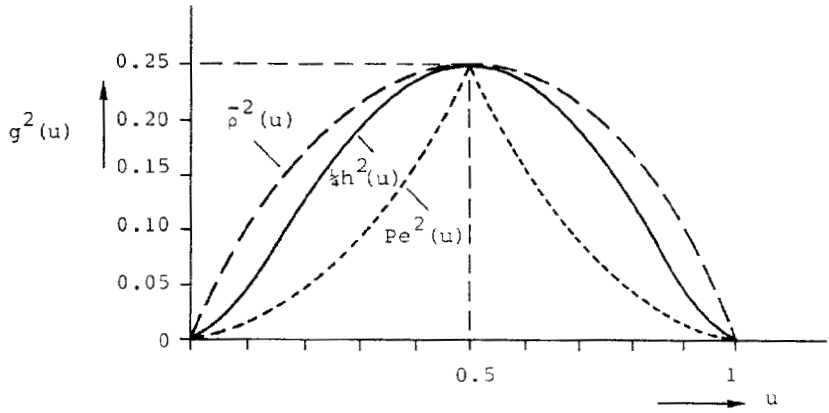


Figure 4.3. The variance weighting for $\bar{\rho}$, H and Pe.

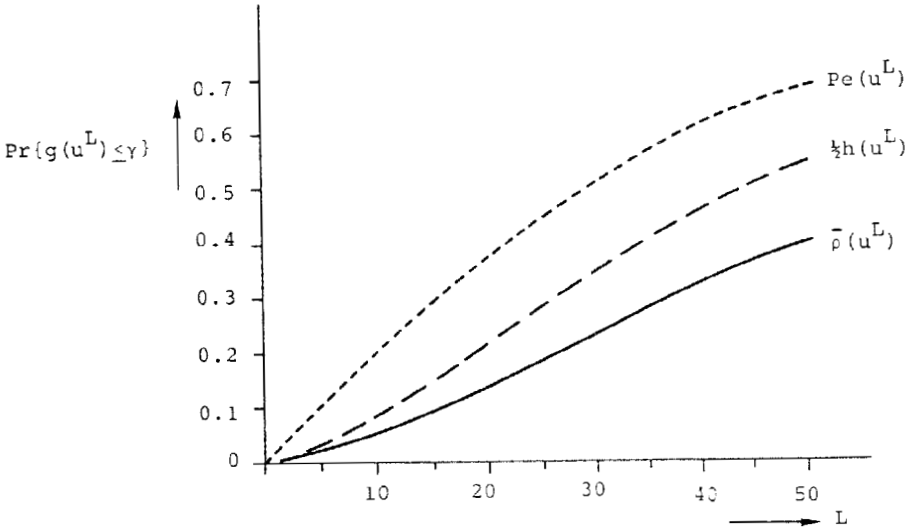


Figure 4.4. Lower bounds (Lemma 4.1) on $Pr\{g(u^L) \leq \gamma\}$ for the different measures in a memoryless SSC-model with p = 0.6.

Figure 4.4. in combination with figure 4.1 tells us that $Pe(K/E^L)$ for a given $\gamma$ is more reliable than the other measures. Finally it is observed that in the SSC-model $Pe(K/E^L)$ at L=UD (UD in RC-model) is almost constant for different values of p and is approximately 0.12 even for UD = $10^6$.

From the SSC-model it seems that Pe is a good and natural measure of theoretical security. For this reason we shall briefly pay attention to its behaviour inan RC-model.

The next theorem is a direct consequence of Hellman's definition of an RC-model and the expected number of spurious key decipherments [3].

Theorem 4.1. The average probability of error (or probability of incorrect key identification) in a random cipher model is given by

$$Pe_{RC}(K/E^L) = \frac{|K| - 1}{|K|} \cdot 2^{-L.R} \approx 2^{-L.R},$$

where

$$R = \log |M| \cdot \left(1 - \frac{H_L(M)}{\log |M|}\right).$$

Proof. There are $|K|$ different and independent keys so that $Pe_{RC}(K/E^L) = \frac{\bar{n}_k}{|K|}$ in which $\bar{n}_k$ is the average number of spurious key decipherments. According to Hellman [3, theorem 1] we have $\bar{n}_k = (|K| - 1).2^{-L.R}$ with $R = \log |M| - H_L(M)$. Substitution yields the theorem. If the key space is sufficiently large we have the nice approximation $2^{-L.R}$.

□

In a similar way the other theorems in [3] can be adapted in terms of Pe too.

Remark. It is important that the assumptions imposed by the RC-model be reasonable for the real secrecy system including the language used. For example, not only the uniformly distributed assumption must be considered but also the effective size of the key space which depends highly on the language used and on the length of the intercepted text. For large L the dependence may be negligible, but for small and moderate values one has to face the fact that some of the keys act similarly, i.e. key residue classes must be considered instead of the single keys. If a key residue class is detected with a small probability of error the remaining keys in this class are indistinguishable. At best one can choose a key according to an arbitrary rule. This introduces an extra error which depends on the size of the residue class. Note that data compression reduces this extra error. So when one's aim is to protect the key, data compression must be considered with care.

At unicity distance it holds that $H(K) = L.R$. Now the next corollary follows immediately from theorem 4.1 and the corresponding remark.

Corollary 4.1. The average probability of error (or probability of incorrect key identification) in a random cipher model at unicity distance is given by

$$Pe_{RC}(K/E^{UD}) = \frac{|K|-1}{2|K|} \approx \frac{1}{2|K|} .$$

□

Note that $Pe_{RC}(K/E^{UD}) = 0.25$ for $|K| = 2$. For the SSC-model we have found that at L=UD (UD in RC-model) $Pe_{SSC}(K/E^{UD}) \approx 0.12$, which was fairly constant even for a $UD=10^6$. This discrepancy is due to the fact that $Pe_{RC}(K/E^{L})$ is an upper bound on $Pe_{SSC}(K/E^{L})$ and is tight for $L >> UD$.

Example 4.5. In an SSC-model using the English language for small and moderate values of L the effective number of keys is less than 26! This is caused by the fact that the average number of different letters that occur in messages of length L is less than 26. This is illustrated in table 4.1. At UD in an RC-model the average number of different letters per message is about 14. Therefore the average probability of error becomes

$$Pe_{RC}(K/E^{UD}) \approx \frac{12!}{26!} \approx 1.10^{-18}.$$

This means that on the average 1 key residue class to every $10^{18}$ key residue classes will be incorrectly identified from the effective number of keys induced by the ciphertext of UD length. The actual $Pe_{RC}(K/E^{UD})$ depends on the size of the key residue class too, which may be rather large. Nevertheless when we know the key residue class we know the message too. This explains why it is almost always possible to get a unique solution at UD.

□

As stated in corollary 4.1 the UD in an RC model defines a Pe which depends on the size of the key space (the larger the size of the key space, the smaller Pe). As a result the meaning of the UD for different sizes of the key space is also different, in the sense of Pe. Actually that is not what one prefers. It is desirable to have a UD for which the explanation is independent of the size of the key space. From the above arguments it seems that linking the UD to Pe leads to a better and more adequate explanation of the UD. For this reason we will generalize the concept of UD in terms of Pe and call the new distance the Pe-security distance (Pe-SD).

| Message length L (Characters) | Average number of different letters per message |
|---|---|
| 5 | 4.5 |
| 10 | 7.8 |
| 15 | 10.2 |
| 20 | 12.0 |
| 25 | 13.4 |
| 30 | 14.5 |
| 40 | 16.1 |
| 50 | 17.3 |
| 75 | 19.2 |
| 100 | 20.4 |
| 200 | 22.4 |
| 300 | 23.0 |
| 400 | 23.4 |
| 500 | 23.7 |
| 700 | 24.2 |
| 1000 | 24.6 |
| 1500 | 25.2 |

Table 4.1. The average number of different letters in L letters of English text. This table was adapted from Meyer and Matyas [19, table 12.3]

Definition 4.2. The Pe-security distance is defined by

$$L_m(\gamma) = \underset{L}{MIN} \{L \in \mathbb{R}^+ | Pe_m(K/E^L) \leq \gamma\},$$

where

m is the actual cipher model and

$\gamma$ is a value of Pe. □

Remark. Depending on what one's object is (the key or the message), the Pe-security distance (for the N-ary case) can be based on $Pe_m(K/E^L)$ or on $Pe_m(M/E^L)$. From the definition it follows that the Pe-SD depends on the model "m" used (including the source) and the desirable value of Pe "$\gamma$". The average performance of the Pe-SD is natural and clear.

Corollary 4.2. The Pe-security distance includes the original unicity distance in a random cipher model as a special case.

Proof. After substitution of

$$Pe_{RC}(K/E^L) = \frac{|K| - 1}{|K|} \cdot 2^{-L.R} \text{ and } \gamma = \frac{|K| - 1}{|K|^2},$$

with $R = \log |M| \cdot (1 - \frac{H_L(M)}{\log |M|})$, one easily obtains

$$\underset{L}{\text{MIN}} \{L \in \mathbb{R}^+ | L \geq \frac{\log |K|}{R} \},$$

which is the original UD inan RC-model. ∎

For the SSC-model with redundancy R the Pe-SD characteristics are given in figure 4.5 for different values of $\gamma$. Note that Pe at UD is almost constant, in conformity with the predictions from the RC-model.

If determining $L_m(\gamma)$ in a direct manner is quite involved one can make use of the lower bounds given in the previous sections.

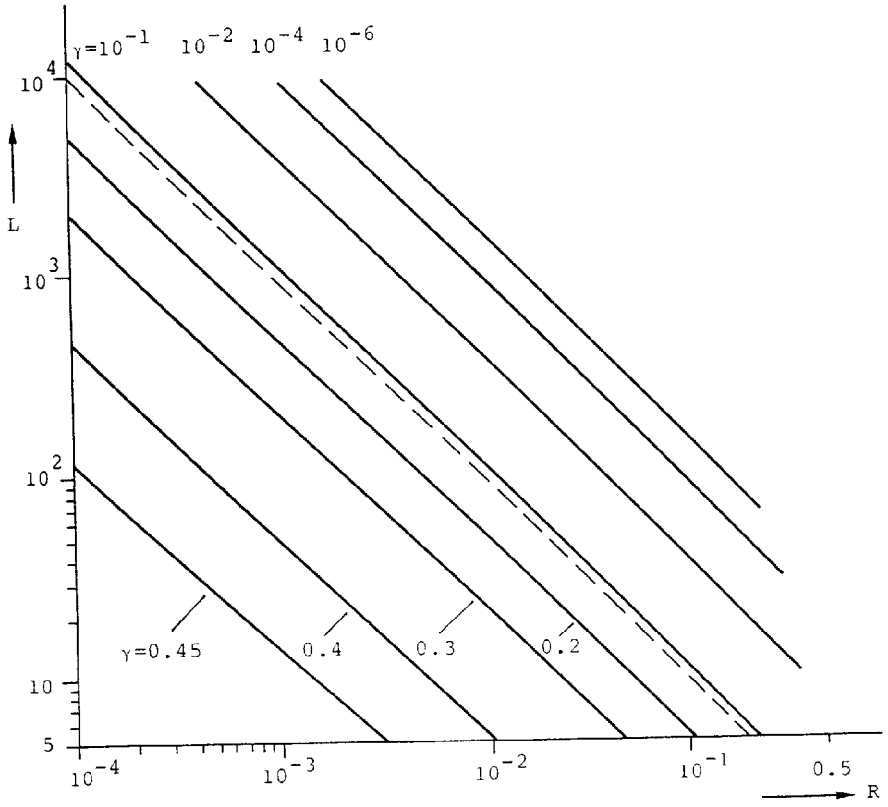


Figure 4.5. The Pe-SD characteristics for an SSC using a binary memoryless source with redundancy R. The dotted line represents Pe at UD.

Example 4.6. For a PC-model using a discrete memoryless source with a priori probabilities p and q we have for the Bhattacharyya coefficient

$$L_{PC}(\gamma) \geq \underset{L}{\text{MIN}} \{L \in \mathbb{R}^+ \mid \tfrac{1}{2} \cdot (1 - \sqrt{1 - (4pq)^L}) \leq \gamma\},$$

from which it is easily found that

$$L_{PC}(\gamma) \geq \frac{\log [1 - (1-2\gamma)^2]}{\log(4pq)} .$$

□

The Pe-SD can be applied in the reverse direction too, i.e. for a given L the corresponding expected value of $\gamma$ can be found. Using the same arguments lower bounds on Pe can be considered to determine $\gamma$.

Example 4.7. Again, consider an SSC using a discrete memoryless source with p=0.7 and L=7 [2, example 7.2]. Jürgensen and Matthews stated that this system is highly insecure even though $H(K/E^7) \approx 0.44$ is fairly large. Since $H(K/E^L)$ itself defines an upper bound on $Pe(K/E^L)$, one must make use of Fano's inequality $H(K/E^L) \leq H(Pe) + Pe.\log(N-1)$. From N=2 and $H(K/E^7) \approx 0.44$ it is found that $\gamma \geq 0.09$. Therefore we may conclude that the system for the given source is indeed insecure.

□

It is illustrated by the example why the key equivocation itself, judged as measure of theoretical security, behaves poorly: it is an upper bound and usually only tight for large L. Although the key equivocation may be a poor measure of security in many cases, it certainly does not degrade the use of Shannon's information measure in cryptanalysis. The strength of this measure can be explained by the natural interpretation and accordingly by the convenient way of manipulating between different pd's. For example, this has been demonstrated by Lu [8].

References

[1] C.E. Shannon, Communication theory of secrecy systems, Bell Syst. Tech. J. 28, pp. 656-715, (1948).

[2] H. Jürgensen and D.E. Matthews, Some results on the information theoretic analysis of cryptosystems, Proc. of CRYPTO'83, Santa Barbara, California, August 1983, pp. 303-356.

[3] M.E. Hellman, An extension of the Shannon Theory Approach to Cryptography, IEEE Trans. Inform. Theory IT-23, pp. 289-294 (1977).

[4] R. Blom, Bounds on Key Equivocation for Simple Substitution Ciphers, IEEE Trans. Inform. Theory IT-25, pp. 8-18 (1979).

[5] R. Blom, An Upper Bound on the Key Equivocation for Pure Ciphers, IEEE Trans. Inform. Theory IT-30, pp. 82-84 (1984).

[6] J.G. Dunham, Bounds on Message Equivocation for Simple Substitution Ciphers, IEEE Trans. Inform. Theory IT-26, pp. 522-527 (1980).

[7] A. Sgarro, Error Probabilities for Simple Substitution Ciphers, IEEE Trans. Inform. Theory IT-29, pp. 190-198 (1983).

[8] S.C. Lu, The Existence of Good Cryptosystems for Key Rates Greater than the Message Redundancy, IEEE Trans. Inform. Theory IT-25, pp. 475-477 (1979).

[9] L. Kanal, Patterns in pattern recognition: 1968-1974, IEEE Trans. Inform. Theory IT-20, pp. 697-722 (1974).

[10] C.H. Chen, Statistical pattern recognition, Hayden Book Co., Rochelle Park, New Jersey (1973).

[11] I. Csiszar, Information-type measures of difference of probability distributions and indirect observations, Stud. Sci. Math. Hungary. 2, pp. 299-318 (1967).

[12] D.E. Boekee and J. van Tilburg, Bounds on the Bayesian Error Probability using Concave Functions, to appear.

[13] D.E. Boekee and J.C. Ruitenbeek, A Class of Lower Bounds on the Bayesian Probability of Error, Information Sciences 25, pp. 21-35, (1981).

[14] D.E. Boekee and J.C.A. van der Lubbe, Some Aspects of Error Bounds in Feature Selection, Pattern recognition, Vol. 11, pp. 353-360 (1979).

[15] T. Ito, Approximate Error Bounds in Pattern Recognition, Machine Intelligence, Vol. 7, pp. 369-376, Edinburgh Univ. Press (1972).

[16] R. Blom, On Pure Ciphers, Internal. Rep. LiTH-ISY-I-0286, Linköping University, Sweden (1979).

[17] V.A. Kovalevsky, On the Criteria for the Information Content of a System of Features, In: Image Pattern Recognition, pp. 67-90, (1980).

[18] J. van Tilburg, Decisions and Selections based on the Bayesian Error Probability with Shannon Information, Certainty and f-divergence, Thesis, Delft Univ. of Techn. (1984, in Dutch).

[19] C.H. Meyer and S.M. Matyas, Cryptography: a new dimension in computer data security, Wiley,NY (1982).