Bounds on Key Equivocation for Simple Substitution Ciphers

Rolf Blom

Bounds on Key Equivocation for Simple Substitution Ciphers

ROLF J. BLOM

Abstract—The equivocation of the key for a simple substitution cipher is upper and lower bounded, when the message source is memoryless. The bounds are shown to be exponentially tight. The results are compared with random ciphering. It is observed that the exponential behavior of the equivocation of the key is not determined by the redundancy in the message source, but by the symbol probabilities which are closest in a certain sense.

I. INTRODUCTION

CIPHERS are used to limit the ability of a wiretapper to discover the content of an intercepted message. In [1] Shannon laid down the theoretical framework for analysis of such a situation and introduced a theory of secrecy systems. A secrecy system is defined as a family of uniquely reversible transformations $\mathcal{S} = \{s_{j}(\cdot)\}_{i}^{J}$ of a set of possible messages $\mathcal{M} = \{m_{j}\}_{i}^{N}$ into a set of cryptograms $\mathcal{C} = \{c_{j}\}_{i}^{J}$, the transformations having associated probabilities $\{p_{i}\}_{i}^{J}$. A block diagram depicting the behavior of a secrecy system is shown in Fig. 1. The message source symbols are transformed by the encipherer into cryptogram symbols before they are transmitted over the channel. To recover the message at the receiving end the inverse transformation is performed by the decipherer. The transformation and inverse transformation used are specified by the outcome of the key source.

When evaluating the strength of a secrecy system, it is assumed that the wiretapper knows the set of transformations $\mathcal{S}$ and the statistics of the message and the key sources. Given this information, but not the actual key, the wiretapper tries to estimate the message and/or the key from an intercepted cryptogram. Under these circumstances it is shown in [1, pp. 667–668] that the conditional entropies of the key and of the message given the cryptogram can be used as measures of the strength of the system. The conditional entropies are called the equivocation of the key and of the message, respectively.

In general it is hard to explicitly calculate these equivocations. Therefore, Shannon [1] introduced random ciphers (or random codes), and he and later Hellman [2] analyzed their properties. In [1, p. 698] it is proposed that complex "practical" ciphers behave approximately as random ciphers. On the other hand, it is stated in [2] that random ciphers perform much more poorly than carefully designed ciphers. In this paper we derive an upper bound on the key equivocation for simple substitution ciphers that is exponentially tight. This bound together with calculations of the equivocation are compared with the equivocation of a corresponding random cipher.

In Section II we formally state the problem and give the necessary background. Section III contains the derivation of expressions on the equivocation of the key that are used in Section IV to obtain upper and lower bounds. In Section V the results are discussed and compared with random ciphers.

II. PROBLEM STATEMENT AND PRELIMINARIES

Refer to Fig. 1. The message source is discrete and memoryless with alphabet $\mathcal{M} = \{1, 2, 3, \cdots, N\}$. The probability of a symbol $n$ is $P_{M}(n) = q_{n}$. The cryptogram alphabet $\mathcal{C}$ is taken to be the same as $\mathcal{M}$. The set of transformations $\mathcal{S} = \{s_{j}(\cdot)\}_{i}^{J}$ is the set of all invertible transformations of $\mathcal{M}$ onto $\mathcal{C}$. Thus the number of elements in $\mathcal{S}$ is $J = N!$. The key and the message sources are independent, and the keys are equiprobable, i.e., $P_{K}(j) = 1/N!$. We will refer to the cipher defined by $\mathcal{S}$ above as a simple substitution cipher. We note that $\mathcal{S}$ is a group of transformations and that the transformations could be seen as permutations of the message alphabet.
Now a word about notation. Let $S$ be an arbitrary finite set. A sequence of length $L$ of symbols in $S$ will be written as

$$s^L = (s_1, s_2, \cdots, s_L)$$  \hspace{1cm} (1)

where subscripted letters denote the components and superscripted boldface letters denote sequences. The ensemble of all sequences of length $L$ is written $S^L$. A similar convention applies to random sequences and variables which are denoted by uppercase letters.

A transformation of a message symbol $m \in S$ will be written as

$$t_j(m) = e,$$  \hspace{1cm} (2)

and we will use the same notation for transformations of a sequence of message symbols

$$t_j(m^L) = (t_j(m_1), t_j(m_2), \cdots, t_j(m_L)) = e^L$$  \hspace{1cm} (3)

which should not cause any confusion. We also define $t_\circ(\cdot) \in S$ to be the identity transformation. The notation of standard information quantities are as defined by Gallager [3], and the wiretapper’s equivocation of the key is written $H(K | E^L)$. The logarithms involved in this paper are taken to the base $e$. Hence entropies and equivocations will be expressed in nats/symbol.

The main object of this paper is to find exponentially tight bounds on the equivocation of the key. However, before doing that we first derive a general lower bound on $H(K | E^L)$ without using the assumption that the message source is memoryless. Then we make an observation about the general behavior of $H(K | E^L)$ when the message source is memoryless.

The lower bound can be obtained by writing

$$H(K | E^L) = H(K | E^L) - H(E^L)$$  \hspace{1cm} (4)

and using the equalities

$$H(K | E^L) = H(K | M^L) = H(K) + H(M^L).$$  \hspace{1cm} (5)

The first equality in (5) is due to the fact that knowing $K$ and $E^L$ is equivalent to knowing $K$ and $M^L$, because all $t_j \in S$ are invertible. The second equality follows from the independence of the message and key sources. Combining (4) and (5) gives

$$H(K | E^L) = H(K) + H(M^L) - H(E^L)$$  \hspace{1cm} (6)

which also can be found in [1, p. 687]. There are $N$ symbols in both $S$ and $S^L$. Thus we can upperbound $H(E^L)$ as

$$H(E^L) \leq L \cdot \log(N)$$  \hspace{1cm} (7)

and write the redundancy $D_L$ of $L$ message characters as

$$D_L = L \cdot \log(N) - H(M^L).$$  \hspace{1cm} (8)

Combining (6), (7), and (8) gives the lower bound

$$H(K | E^L) \geq H(K) - D_L.$$  \hspace{1cm} (9)

The fundamental nature of this lower bound leads us to state this result as a theorem.

**Theorem 1:** If the key and message sources are independent, the key equivocation of a secrecy system is lower bounded by (9).

When the message source is memoryless, (9) can be written as

$$H(K | E^L) \geq H(K) - L \cdot \log(N) - H(M).$$  \hspace{1cm} (10)

We observe that (10) is equal to the approximate expression for the key equivocation of a random cipher [1, pp. 691–693] when

$$L < U = H(K) / [\log(N) - H(M)].$$  \hspace{1cm} (11)

$U$ is called the unicity distance. The interpretation is that after the interception of $U$ symbols, it is almost always possible to get a unique solution to a random cipher. We see that up to the point when the random cipher becomes uniquely solvable, the key equivocation of the cipher behaves as the general lower bound in (10). Thus the above is a simpler and more general derivation of Hellman’s result [2] that a random cipher is essentially the worst possible.

From the properties of conditional entropy, it is evident that $H(K | E^L)$ is monotonically decreasing with $L$. When the message source is memoryless, the equivocation of the key is also convex in the sense that

$$H(K | E^L) - H(K | E^{L+1}) \geq H(K | E^{L+1}) - H(K | E^{L+2}).$$  \hspace{1cm} (12)

To see this, subtract the right side of (12) from the left side, and substitute (6). Then we get

$$H(K | E^L) - 2H(K | E^{L+1}) - H(K | E^{L+2}) = H(K) + L \cdot H(M) - H(E^L)
= 2[H(K) + (L+1) \cdot H(M) - H(E^{L+1})]
+ H(K) + (L+2) \cdot H(M) - H(E^{L+2})$$
$$= H(E^{L+1}) - H(E^L) - [H(E^{L+2}) - H(E^{L+1})]$$
$$= H(E_{L+1} | E_1 E_2 \cdots E_L) - H(E_{L+2} | E_1 E_2 \cdots E_{L+1})$$
$$> H(E_{L+1} | E_1 E_2 \cdots E_L) - H(E_{L+2} | E_2 E_3 \cdots E_{L+1}) = 0.$$  \hspace{1cm} (13)

The inequality in (13) is due to the reduction of the number of variables upon which the conditioning is made in the second term. The last expression is zero because of the stationarity of the process.

**III. THE EQUIVOCATION OF THE KEY**

In this section we derive an exact expression for $H(K | E^L)$ in terms of the message symbol probabilities. This expression is used to calculate exact values of the key.
equivocation to which the bounds can be compared. It is also used as a starting point in the derivation of an upper bound of $H(K|E^L)$ when the message source is binary.

To obtain the desired expression for $H(K|E^L)$, we write

$$H(K|E^L) = \sum_{k=1}^{N^L} \sum_{e^L \in E^L} P_{E^L|K}(e^L, k) \log \left( \frac{\sum_{l=1}^{N^L} P_{E^L|K}(e^L, l)}{P_{E^L|K}(e^L, k)} \right). \quad (14)$$

But

$$P_{E^L|K}(e^L, k) = \sum_{m^L \in \mathbb{N}^L} P_{E^L|KM^L}(e^L|m^L) P_K(k) P_{M^L}(m^L), \quad (15)$$

and

$$P_{E^L|KM^L}(e^L|m^L) = \begin{cases} 1, & t_k(m^L) = e^L \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

because $t_k$ is deterministic and invertible. Hence (15) can be written as

$$P_{E^L|K}(e^L, k) = \frac{1}{N!} P_{M^L}(t_k^{-1}(e^L)) \quad (17)$$

where we use the assumption that all keys are equiprobable. To proceed we introduce a vector $y=(y_1,y_2,\cdots,y_N)$ that contains the frequencies of the different symbols in the cryptogram $e^L$, that is, $y_n$ is the number of times the symbol $n$ appears in the cryptogram. Let $x=(x_1,x_2,\cdots,x_N)$ contain the corresponding frequencies of $m^L=t_k^{-1}(e^L)$. Then $x$ is a permutation of $y$ and the components of $x$ and $y$ satisfy the relation

$$x_n = y_{t_k(n)}. \quad (18)$$

The message source is memoryless, and we get

$$P_{E^L|K}(e^L, k) = \frac{1}{N!} q_1^n q_2^n \cdots q_N^n = \frac{1}{N!} \prod_{n=1}^{N} q_{t_k(n)}^n. \quad (19)$$

We also observe that the following equalities are true:

$$N! \left( \sum_{j=1}^{N} P_{E^L|K}(e^L, j) \right) = \sum_{j=1}^{N} \prod_{n=1}^{N} q_{t_k(n)}^n = \sum_{n=1}^{N} \prod_{n=1}^{N} q_{t_k(n)}^n. \quad (20)$$

To see this, note that the summation is done over all permutations of the indices of either the exponents or the exponentiated factors.

The sum over all cryptograms in $\mathbb{S}^L$ in (14) can now be expressed as a sum over all frequency vectors $y$ for which $|y| = \sum_{n=1}^{N} y_n = L. \quad (21)$

Hence, after substitution of (19) into (14), the equalities in (20) with $x$ defined by (18) give

$$H(K|E^L) = \sum_{k=1}^{N^L} \frac{1}{N!} \sum_{|y|=L} \frac{L!}{|y|!} \prod_{n=1}^{N} q_{t_k(n)}^n \log \left( \prod_{n=1}^{N} q_{t_k(n)}^n \right) \prod_{n=1}^{N} \frac{q_{t_k(n)}^n}{\prod_{n=1}^{N} q_{t_k(n)}^n}$$

$$= \sum_{n=1}^{N} \prod_{n=1}^{N} \frac{q_{t_k(n)}^n}{\prod_{n=1}^{N} q_{t_k(n)}^n}$$

$$= \sum_{|x|=L} \prod_{n=1}^{N} q_{t_k(n)}^n \log \left( \prod_{n=1}^{N} q_{t_k(n)}^n \right) \prod_{n=1}^{N} \frac{q_{t_k(n)}^n}{\prod_{n=1}^{N} q_{t_k(n)}^n} \quad (22)$$

which is the desired result.

IV. UPPER AND LOWER BOUNDS

To obtain the upper bound we have to prove three inequalities related to entropy functions. We state these inequalities in a general setting in the three lemmas below. For proofs see Appendix A.

Lemma 1: If

$$\sum_{j=1}^{J} p_j = 1, \quad p_j > 0,$$

then

$$- \sum_{j=1}^{J} p_j \log (p_j) < \log \left( - \sum_{j=1}^{J} \sum_{j=1}^{J} \sqrt{p_j p_j} \right). \quad (23)$$

Lemma 2: If

$$\sum_{j=1}^{J} p_{ij} = p_i, \quad \sum_{i=1}^{I} p_i = 1, \quad p_{ij} > 0,$$

then

$$\sum_{j=1}^{J} \sum_{i=1}^{I} p_{ij} \log \frac{p_{ij}}{p_{ij}} < \log \left( - \sum_{j=1}^{J} \sum_{j=1}^{J} \sum_{j=1}^{J} \sqrt{p_{ij} p_{ij}} \right). \quad (24)$$

In the third lemma we improve the bound of (24) for the special case $J=2$, for all $i$.

Lemma 3: If

$$p_{i1} + p_{i2} = p_i, \quad \sum_{i=1}^{I} p_i = 1, \quad p_{i1}, p_{i2} > 0,$$

then

$$\sum_{i=1}^{I} \sum_{j=1}^{2} p_{ij} \log \left( \frac{p_{ij}}{p_{ij}} \right) < 2 \log (2) \sum_{i=1}^{I} \sqrt{p_{i1} p_{i2}}. \quad (25)$$
As a corollary to Lemma 3, we state a simple upper bound to the entropy of a binary source \(I = 1\). Fig. 2 is a plot of this bound.

**Corollary 1**: If a binary source has \(P(1) = p\) and \(P(0) = 1 - p\), we have

\[
 h(p) = -p \log (p) - (1 - p) \log (1 - p) \leq \sqrt{4p(1-p)} \log 2. \tag{26}
\]

It is now possible to derive an upper bound on the equivocation of the key. The bound is given in the following theorem.

**Theorem 2**: If a discrete memoryless source is encrypted by a simple substitution cipher with equiprobable keys and the source alphabet has \(N\) letters with probabilities \(\{q_n\}_{n=1}^N\), we have

\[
 a) \quad H(K|E^L) < \log \left[ 1 + \sum_{l=2}^{N^I} \left( \sum_{n=1}^N \sqrt{q_n q_n(n)} \right)^L \right], \quad N > 2 \tag{27}
\]

\[
 b) \quad H(K|E^L) < \sqrt{4q_1q_2} L \log 2, \quad N = 2. \tag{28}
\]

**Proof:**

a) Applying Lemma 2 to (14) gives

\[
 H(K|E^L) = \sum_{e^I \in \mathcal{E}^I} \left( \sum_{k=1}^{N^I} P_{E^I|K}(e^I, k) \log \left[ \sum_{l=1}^{L^I} P_{E^I|K}(e^I, l) \right] \right) \leq \log \left( \sum_{e^I \in \mathcal{E}^I} \left( \sum_{k=1}^{N^I} \sum_{l=1}^{L^I} P_{E^I|K}(e^I, l) \right) \right). \tag{29}
\]

Using the notation of Section III, substituting (19) into (29), and using an equality similar to (20), we have

\[
 H(K|E^L) < \log \left( \prod_{n=1}^N q_n^{e_n} \right) \left( \prod_{n=1}^N q_n^{e_n} \right) \leq \log \left( \sum_{k=1}^{N^I} \left( \sum_{n=1}^N \sqrt{q_n q_n(n)} \right)^L \right). \tag{30}
\]

However,

\[
 \sum_{k=1}^{N^I} \left( \sum_{n=1}^N \sqrt{q_n q_n(n)} \right) \leq \sum_{k=1}^{N^I} \left( \sum_{n=1}^N q_n q_n(n) \right)^L \tag{31}
\]

because the summation over \(l\) is over all permutations of the indices. The right side of (31) is independent of \(k\). Thus substitution of (31) in (30) and summation over \(k\) give the upper bound in (27).

b) When \(N = 2\), (22) reduces to

\[
 H(K|E^L) = \sum_{x=0}^{L^I} \left( L^I x \right) \left( 1^{L^I-x} q_1^{L^I-x} + q_1^{L^I-x} q_1^{L^I-x} \right) \leq \sum_{x=0}^{[L^I/2]} \left( L^I x \right) R(L, x) \left( q_1^{L^I-x} q_1^{L^I-x} + q_1^{L^I-x} q_1^{L^I-x} \right) + \left( L^I - x \right) \left( q_1^{L^I-x} q_1^{L^I-x} + q_1^{L^I-x} q_1^{L^I-x} \right) \tag{32}
\]
where $[L/2]$ is the largest integer less than or equal to $L/2$ and
\[
R(L,x) = \begin{cases} 1, & x \neq L/2 \\ 1/2, & x = L/2. \end{cases} \tag{33}
\]

It is now easy to apply Lemma 3 to (32), which gives
\[
H(K|E^L) \leq \log \left( \sum_{x=0}^{[L/2]} \left( \frac{L}{x} \right) R(L,x) \right)
= \log \left( \frac{L!}{L - x} \right) \sum_{x=0}^{[L/2]} \left( \frac{L}{x} \right) R(L,x)
= \sqrt{4q_1q_2} \log \left( \frac{L!}{L} \right) \log (2). \tag{34}
\]

Remarks: If we let
\[
a_i = \sum_{n=1}^{N} \sqrt{q_n} q_{i(n)}, \tag{35}
\]
we can write (27) as
\[
H(K|E^L) \leq \log \left( 1 + \sum_{i=2}^{N} a_i^2 \right). \tag{36}
\]

Cauchy’s inequality shows that $a_i < 1$, $i = 1, 2, \ldots, N$, because
\[
a_i^2 = \left( \sum_{n=1}^{N} \sqrt{q_n} \sqrt{q_{i(n)}} \right)^2 \leq \left( \sum_{n=1}^{N} q_n \right) \left( \sum_{n=1}^{N} q_{i(n)} \right) = 1. \tag{37}
\]

A necessary and sufficient condition for equality in (37) is that
\[
q_n = q_{i(n)}, \quad n = 1, 2, \ldots, N. \tag{38}
\]

When all $q_n$ are distinct, (38) is true only for $i = 1$, and the bound will go to zero when $L$ goes to infinity. But if some $q_n$ are equal, (38) will be true for additional values of $i$. To find the limiting values of the bound in (36) and of $H(K|E^L)$ in such a case, assume that all $q_n$ are equal to one of $N_1 < N$ different values $\{q^*_n\}_1^{N_1}$, and define sets $\mathcal{G}_n$ as
\[
\mathcal{G}_n = \{ i | q_i = q^*_n, i \in \{1, 2, \ldots, N\} \}, \quad n = 1, 2, \ldots, N_1. \tag{39}
\]

Let $\mathcal{E}_1$ be a set of indices defined by
\[
\mathcal{E}_1 = \{ i | q_i, t_i(i) \in \mathcal{G}_n, \text{ for all } i = 1, 2, \ldots, N \}. \tag{40}
\]

Let $u_i$ be the number of elements in $\mathcal{G}_n$. Then there are $u_i!$ invertible transformations of $\mathcal{G}_n$ onto $\mathcal{G}_n$. Hence the number of elements $d$ in $\mathcal{E}_1$ is
\[
d = \prod_{n=1}^{N_1} u_i! \tag{41}
\]

We also observe that the set $\mathcal{G}_1$ of transformations
\[
\mathcal{G}_1 = \{ t_i(\cdot)|l \in \mathcal{E}_1 \} \tag{42}
\]
is a subgroup of the group generated by the elements of $\mathcal{G}$. This subgroup generates a coset partitioning of $\mathcal{G}$, and we see that if $t_1(\cdot)$ and $t_2(\cdot)$ both belong to the same coset, then
\[
q_{i(n)} = q_{i(n)}, \quad n = 1, 2, \ldots, N. \tag{43}
\]

The number of elements in each coset is $d$, and the number of cosets is $N!/d$. We can use this fact by defining a new set $\mathcal{E}_2$ of indices such that the set
\[
\mathcal{E}_2 = \{ t_1(l)|l \in \mathcal{E}_2 \} \tag{44}
\]
contains one element from each coset. We assume that $t_1(\cdot)$ represents the coset formed by $\mathcal{E}_2$. Hence $l \in \mathcal{E}_2$, and for notational reasons let us define $\mathcal{E}_3 = \mathcal{E}_2 \setminus \{l\}$. Then the upper bound in (27) can be written as
\[
H(K|E^L) \leq \log \left( \sum_{i=1}^{N_1} \left( \sum_{n=1}^{N} \sqrt{q_n} q_{i(n)} \right)^{L_i} \right)
= \log \left( d \sum_{i=1}^{N_1} \left( \sum_{n=1}^{N} \sqrt{q_n} q_{i(n)} \right)^{L_i} \right)
= \log (d) + \log \left( 1 + \sum_{i \in \mathcal{E}_3} a_i^2 \right). \tag{45}
\]

From the definitions of $\mathcal{G}_1$ and $\mathcal{E}_2$, it is obvious that $a_1 < 1$ when $l \in \mathcal{E}_2$, and consequently (45) shows that the limit of the upper bound is $\log (d)$ when $L$ goes to infinity.

We can also show that $H(K|E^L) > \log (d)$. To see this use (22) to write
\[
H(K|E^L) = \sum_{|s|=L} \frac{L!}{x_1!x_2! \cdots x_N!} \prod_{n=1}^{N} q_n^{s_n} \log \left( \frac{\prod_{n=1}^{N} q_n^{s_n}}{\prod_{n=1}^{N} q_n^{s_n}} \right)
= \log (d) + \sum_{|s|=L} \frac{L!}{x_1!x_2! \cdots x_N!} \prod_{n=1}^{N} q_n^{s_n} \log \left( \frac{\prod_{n=1}^{N} q_n^{s_n}}{\prod_{n=1}^{N} q_n^{s_n}} \right)
> \log (d). \tag{46}
\]

Then (45) and (46) show that both the upper bound in (27) and $H(K|E^L)$ have the same limiting value when $L$ goes to infinity. From (36) it is obvious that the bound has the correct value $\log (N!)$ at $L=0$.

Fig. 3 shows two examples of the bound when $N=4$. In this figure, as in the following ones, the parameters of the plot are found in the upper right corner. $N$, $L$, and $S$ denote the number of symbols in the message source, the maximum $L$, and the stepsize in $L$ used in calculating $H(K|E^L)$, respectively. $H$ and $U$ are the entropy of the
Fig. 3. Two examples of bound on $H(K|E^v)$ when $N=4$. 

(a) 

(b) 

N = 4  
$q_1 = 0.4$  
$L = 100$  
$q_2 = 0.3$  
$S = 2$  
$q_3 = 0.2$  
$H = 1.28$  
$U = 29.86$  

N = 4  
$q_1 = 0.31$  
$L = 100$  
$q_2 = 0.27$  
$S = 2$  
$q_3 = 0.24$  
$H = 1.37$  
$U = 171.45$
message source and the unicity distance, $q_1, q_2, \cdots$ are the symbol probabilities of the message source.

It is possible to show that the bounds in Theorem 2 are exponentially tight. To do this we start by finding a new lower bound for the case $N = 2$ and $L$ even.

**Theorem 3:** If a binary memoryless source is enciphered by a simple substitution cipher with equiprobable keys and the symbol probabilities of the message source are $q_1, q_2$, we have

$$H(K|E^L) \geq \frac{1}{\sqrt{L}} A(L) \sqrt{4q_1 q_2} \log (2)$$

(47)

$$A(L) = \sqrt{\frac{2}{\pi}} \left( \frac{1}{1 + \frac{1}{2L}} \right)^2$$

(48)

for $L = 2, 4, 6, \cdots$.

**Proof:** We start with the expression (32) for the equivocation used in the proof of Theorem 2:

$$H(K|E^L) = \sum_{x=0}^{L} \binom{L}{x} q_1^x q_2^{L-x} \log \left( \frac{q_1 q_2^{L-x} + q_1^{L-x} q_2}{q_1 q_2} \right).$$

(49)

As a lower bound we take the term for $x = L/2$ in (49)

$$H(K|E^L) \geq \left( \frac{L}{L/2} \right)^2 q_1 q_2^{-L/2} \log (2)$$

(50)

Lower bounding the binomial coefficient in (50) by Stirling’s formula gives

$$\binom{L}{L/2} \geq \sqrt{\frac{2}{\pi}} \left( \frac{1}{1 + \frac{1}{2L}} \right)^2.$$ 

(51)

Substitution of (51) into (50) and identification of terms proves the theorem.

Comparing (28) in Theorem 2 and (47) makes it obvious that we have exponentially tight bounds on the equivocation when $N = 2$. Fig. 4 shows the bounds for two different cases.

To get a lower bound that holds for all values of $L$, we observe that

$$\frac{1}{\sqrt{L}} A(L) \geq \frac{1}{\sqrt{L+1}} A(L+1), \text{ for } L \geq 2$$

(52)

$$H(K|E^L) \geq H(K|E^{L+1}).$$

(53)

Then when $L \geq 1$,

$$H(K|E^L) \geq \frac{1}{\sqrt{L+1}} A(L+1) \sqrt{4q_1 q_2} L^{L+1} \log (2).$$

(54)

Finally by evaluation of (54) for $L = 0$, we obtain

$$H(K) = \log (2) > \sqrt{\frac{2}{\pi}} \frac{4}{9} \log (2).$$

(55)

Hence we have proved the following corollary:

**Corollary 2:** If a binary memoryless source is enciphered by a simple substitution cipher with equiprobable keys and the symbol probabilities of the message source are $q_1, q_2$, we have

$$H(K|E^L) \geq \frac{1}{\sqrt{L+1}} B(L) \sqrt{4q_1 q_2} L^{L+1} \log (2)$$

(56)

$$B(L) = \sqrt{\frac{2}{\pi}} \left( \frac{1}{1 + \frac{2}{2L+1}} \right)^2.$$ 

(57)

Now we show that (27) in Theorem 2 which applies for $N > 2$ is exponentially tight. To reach our goal we start by simplifying the upper bound as it is stated in (45) by using the standard inequality $\log (1 + x) \leq x:

$$H(K|E^L) < \log (d) + \log \left( 1 + \sum_{j \in \mathbb{L}_2} a_j \right)$$

(58)

where

$$a_i = \max \{a_i\}.$$ 

(59)

To determine $t_i(\cdot)$, we write

$$a_i = \sum_{n=1}^{N} \sqrt{q_n q_{i(n)}} = 1 - \frac{1}{2} \sum_{n=1}^{N} \left( \sqrt{q_n} - \sqrt{q_{i(n)}} \right)^2.$$

(60)

Let $b_{ij}$ be defined by

$$b_{ij} = |\sqrt{q_i} - \sqrt{q_j}|, \quad i, j \in \{1, 2, \cdots, N\},$$

(61)

and let $i = r$ and $j = s$ be the indices for which $b_{ij}$ has its least value greater than zero. We also observe that if

$$q_n \neq q_{i(n)},$$

(62)

for a particular value of $n$, then there must exist another value of $n$ for which (62) is true. Furthermore, because $b_{ij} = b_{ji}$, a transformation yielding the maximum $a_i$ would belong to the coset generated by

$$t(n) = \begin{cases} r, & n = \mu \\ \mu, & n = \nu \\ n, & \text{otherwise}. \end{cases}$$

(63)

Hence we may assume that $t_0$ gives the transformation specified by (63).

Using the form of $H(K|E^L)$ given by the first equality of (46) as a starting point, we get the following lower
Fig. 4. Two examples of bound on $H(K|E^I)$ when $N=2$. 
bound:

\[ H(K|E^L) = \log(d) + \sum_{|x|=L} \frac{L!}{x_1!x_2! \cdots x_N!} \cdot \prod_{n=1}^{N} q_{n}^{x_n} \log \left( \frac{\sum_{j \in \mathcal{E}, \pi = n} L_{\pi}^{x_n}}{\prod_{n=1}^{N} q_n^{x_n}} \right) \]

\[ \geq \log(d) + \sum_{|x|=L} \frac{L!}{x_1!x_2! \cdots x_N!} \cdot \prod_{n=1}^{N} q_{n}^{x_n} \log \left( \frac{\sum_{j \in \mathcal{E}, \pi = n} L_{\pi}^{x_n}}{\prod_{n=1}^{N} q_n^{x_n}} \right) \]

\[ = \log(d) + \sum_{|x|=L} \frac{L!}{x_1!x_2! \cdots x_N!} \cdot \prod_{n=1}^{N} q_{n}^{x_n} \log \left( \frac{\sum_{j \in \mathcal{E}, \pi = n} L_{\pi}^{x_n}}{\prod_{n=1}^{N} q_n^{x_n}} \right) \]

Equations (58) and (70) show the exponential behavior of the bound.

V. DISCUSSION

As is seen in Figs. 3 and 4, the general behavior of the upper bounds in (27) and (28) are quite similar to the exact \( H(K|E^L) \). For small values of \( N \), (27) can easily be evaluated by a computer. The time to compute the bound in Fig. 3 is negligible while the exact computation of \( H(K|E^L) \) took about 12 hours on an Eclipse computer. With increasing \( N \) the bound grows less attractive to evaluate, because it involves the sum of \( N! \) terms exponentiated to \( L \). However, if one allows a degradation of the bound, this difficulty can be circumvented by upper bounding the sum inside the logarithm. One way to this is shown in Appendix B.

From the derivation of the exponential behavior of the bounds, it becomes evident that the exponential behavior of \( H(K|E^L) \) is determined by the symbol probabilities that are most equal, in the sense that \( |\sqrt{q_{\pi}} - \sqrt{q_{\mu}}| > 0 \) is as small as possible. This stands in sharp contrast to the exponential behavior of a random cipher which is determined by the redundancy in the message source [1, pp. 691–693]. According to (10) the behavior of the equivocation of a random cipher for small \( L \) is also determined by the redundancy. Fig. 5 shows the equivocation of two sources with approximately the same entropy. From the figure it is seen that the equivocations behave differently, and so do the bounds.

ACKNOWLEDGMENT

The author would like to thank Prof. I. Ingemarsson for valuable discussions and comments in the various phases of this work. It is also a pleasure to thank Prof. T. Ericsson and the referees for their very helpful comments on how to improve the original manuscript.

APPENDIX A

Proof of Lemma 1

The proof depends on an inequality between the arithmetic and geometric means. From [4, eq. 2.5.2] we get

\[ \prod_{i=1}^{L} a_i \leq \frac{L}{\sum_{i=1}^{L} a_i} \]

Rewriting the left side in (23) and using (71) gives

\[ - \sum_{i=1}^{L} \rho_i \log(p_i) = 2 \sum_{i=1}^{L} \rho_i \log(p_i^{-1/2}) \]

\[ = 2 \log \left( \prod_{i=1}^{L} (p_i^{-1/2})^{2} \right) \leq 2 \log \left( \prod_{i=1}^{L} \sqrt{p_i} \right) \]

\[ = \log \left( \sum_{i=1}^{L} \sqrt{p_i} \right) \]

which is the desired result.
Fig. 5. Two examples of bound on $H(K|E^L)$ when $N=3$ and message sources have approximately the same entropy.
Proof of Lemma 2

To obtain (24) we rewrite its left side and use Lemma 1:

\[ \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \log \left( \frac{p_{ij}}{p_{ij}} \right) = \sum_{i=1}^{I} p_i \sum_{j=1}^{J} \frac{p_{ij}}{p_i} \log \left( \frac{p_{ij}}{p_i} \right) \leq \sum_{i=1}^{I} p_i \log \left( \sum_{j=1}^{J} \frac{p_{ij} p_{ij}}{p_i} \right). \tag{73} \]

The logarithm is a convex function, and so (73) can be upper-bounded by

\[ \log \left( \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} p_{ij} p_{ij} \right). \tag{74} \]

Proof of Lemma 3

Let \( y_i = \sqrt{p_{1i}/p_{2i}} \). Substitution into the left side of (25) gives the following inequality:

\[ \sum_{i=1}^{I} \sqrt{p_{1i} p_{2i}} \left[ \frac{1}{y_i} \log (1+y_i^2) + y_i \log \left( 1 + \frac{1}{y_i^2} \right) \right] \leq 2 \log (2) \sum_{i=1}^{I} \sqrt{p_{1i} p_{2i}}. \tag{75} \]

It is now sufficient to prove that

\[ f(y_i) = \frac{1}{y_i} \log (1+y_i^2) + y_i \log \left( 1 + \frac{1}{y_i^2} \right) \leq 2 \log 2 \tag{76} \]

when \( 0 < y_i < 1 \), because we can, without loss of generality, assume that \( p_{1i} < p_{2i} \). The derivative of \( f(y) \) is

\[ \frac{d}{dy} f(y) = -\frac{1}{y^2} \left[ \left( 1-y^2 \right) \log (1+y^2) + y^2 \log y^2 \right]. \tag{77} \]

When \( 0 < y < 1 \), we can use the convexity of the logarithm to obtain a lower bound

\[ \frac{d}{dy} f(y) \geq -\frac{1}{y^2} \log \left( (1-y^2)(1+y^2)^2 \right) = 0. \tag{78} \]

Thus the derivative is positive, \( f(0) = 0 \), \( f(1) = 2 \log 2 \), which proves (76).

APPENDIX B

Let \( z \) represent the sum inside the logarithm of (36):

\[ z = 1 + \sum_{i=2}^{N} a_i^2. \tag{79} \]

We wish to find an upper bound on \( z \) that is reasonable to calculate when \( N \) is large. The technique we use is to divide the set of all \( a_i \) into groups and to represent all \( a_i \) in a group by the maximum value of the group. For simplicity let us assume that all \( q_k \) are distinct and that \( q_k > q_{k+1} \). To avoid notational troubles we will only explicitly describe the case when the division is into \( N \) groups. Generalizing this procedure to other numbers of groups should be immediate.

Let us define the partitioning by \( N \) sets \( E_i \) of indices of \( l \). For a fixed \( i \in \{1,2,\ldots,N\} \), let

\[ E_i = \{ l \mid a_i(l) = i \}, \quad j = 1,2,\ldots,N. \tag{80} \]

The number of elements in each \( E_i \) is \( (N-1)! \). To find the maximum of \( a_i \) when \( l \in E_i \), we write

\[ a_i = \sqrt{q_i q_j} + \sum_{n=1}^{N} q_{n} q_{(n)} \tag{81} \]

We observe that the sum in (81) is over the pairwise product of elements from two sequences

\[ \{ \sqrt{q_n} \}_{n=1}^{N} \quad \text{and} \quad \{ \sqrt{q_{n-1}} \}_{n=1}^{N} \]

respectively. The elements in the first sequence decrease with \( N \) while the elements in the second sequence could be arbitrarily ordered. However, in \( E_i \) there exists one \( l \) for every possible ordering. We now use the fact that the maximum of a sum such as the one in (81) is reached when both sequences are similarly ordered, that is when both sequences are either increasing or decreasing [4, p. 262]. Thus there is an effective algorithm for calculating

\[ a_i = \max_{l \in E_i} (a_l). \tag{82} \]

It only remains to take care of the set of \( a_l \) defined by \( E_j \). Because of the assumption that all \( q_k \) are distinct, it is only for \( l = 1 \) that \( a_1 = 1 \) when \( l \in E_j \). We can upper bound all other \( a_l \) in this group with \( q_k \) and write an upper bound of \( z \) as

\[ z < 1 + ((N-1)!-1)a_1^2 + (N-1)! \sum_{n=1}^{N} a_n^2. \tag{83} \]

Let us point out that the tightness of the bound depends on the choice of \( i \) in the definition of \( E_i \). This is because the maximum in each group of \( a_i \) depends on the probabilities \( q_k \).

To make the bound better, the groups \( E_i \) could themselves be further divided. The way to do this is to use the same technique as we used above and introduce subgroups such as the one defined by

\[ E_i = \{ l \mid a_i(k) = l, l \in E_j \}. \tag{84} \]

If this process of dividing existing groups continues, the eventual result will be \( z \). How far to go in this process of dividing into subgroups must be decided by how many terms one can afford in the computation of the bound.

REFERENCES