1979-04-17

AN UPPER BOUND ON THE KEY EQUIVOCATION
FOR PURE CIPHERS

Rolf Blom

INTERNAL REPORT
LiTH-ISY-I-0287

ABSTRACT

An upper bound on key equivocation for a pure cipher applied on a memoryless message source is derived.

CONTENTS

I	Introduction	1
II	The Upper Bound	2
III	An Example	4
IV	Discussion	5
	Appendix	6
	Figures	8
	References	10

I. INTRODUCTION

This correspondence is more or less an addendum to a previous paper [1] by the author that gave upper and lower bounds on the key equivocation of the simple substitution cipher applied on a memoryless source. The result presented here which holds for pure ciphers is an upper bound on the key equivocation similar to that of bound a) Theorem 2 in [1].

As the steps in the derivation of this upper bound are almost the same as the steps in the derivation of bound a) in Theorem 2 in [1] we will omit the proof. A summary of the necessary changes in the derivations in [1] to obtain (2) and (3) are given in an Appendix.

II. THE UPPER BOUND

The notation and assumptions of this correspondence comply as far as possible with those of [1]. However, a brief introduction and the specific assumptions used is given below.

The model used is that of a secrecy system. The message source is memoryless and the message and cryptogram alphabets are equal; $M = E = \{1, 2, \dots, N\}$. The a priori probabilities of the messages are $P_M(n) = q_n$. The set of enciphering transformations $T = \{t_j(\cdot)\}_{j=1}^J$ forms a left coset in the group G of all invertible transformations of M onto M and the keys are equiprobable. According to Theorem 3 in [2] this means that the cipher is pure.

As T is a left coset we may define T as $T = \{g(r_j(\cdot))\}_{j=1}^J$ where $g(\cdot) \in G$ and $R = \{r_j(\cdot)\}_{j=1}^J$ is a subgroup in G . We assume that $t_j(\cdot) = g(r_j(\cdot))$. Then it is obvious that

$$R = \{t_k^{-1}(t_\ell(\cdot))\}_{\ell=1}^J \quad \text{for all } k = 1, 2, \dots, J. \quad (1)$$

The equivocation of the key given that a cryptogram sequence of length L is observed is denoted $H(K|E^L)$. $H(K|E^L)$ is measured in nats and all logarithms used are taken to the base e . For a vector $\underline{x} = (x_1, x_2, \dots, x_N)$, $|\underline{x}|$ is defined by $|\underline{x}| = \sum x_i$.

Under the assumptions made above the exact expression of the key equivocation is

$$H(K|E^L) = \sum_{|\underline{x}|=L} \frac{L!}{x_1! x_2! \dots x_N!} \prod_{n=1}^N q_n^{x_n} \log \left(\frac{\sum_{\ell=1}^J \prod_{n=1}^N q_{r_\ell(n)}^{x_n}}{\prod_{n=1}^N q_n^{x_n}} \right). \quad (2)$$

We observe that (2) only depends on the elements of R and not on T itself. Figure 1 gives an explanation of this fact. Recall that T is assumed to be known by the wiretapper. Hence the wiretapper can determine the group R and a generating element $g_1(\cdot)$ of the coset. Both $g(\cdot)$ and $g_1(\cdot)$ belong to T which implies that $g_1(\cdot)$ can be written as $g_1(\cdot) = g(r_i(\cdot))$ for some i . This gives that y , defined in the figure, is equal to $y = r_i^{-1}(g^{-1}(g(r_k(m)))) = r_i^{-1}(r_k(m))$. Thus the cryptanalysis can just as well start with y and there is no dependence on $g(\cdot)$.

Another way to say this is to first observe that the ciphers with T and R as their sets of enciphering transformations respectively are similar and then observe that (2) also gives the key equivocation of R . The same behaviour is present in the upper bound on $H(K|E^L)$ stated in the following theorem.

Theorem 1: If a discrete memoryless source is enciphered with a pure cipher having T as its set of enciphering transformations and the a priori probabilities of the message source are $P_M(n) = q_n$ then

$$H(K|E^L) \leq \log \left(1 + \sum_{\ell=2}^J \left(\sum_{n=1}^N \sqrt{q_n q_{r_\ell(n)}} \right)^L \right) \quad (3)$$

where $r_\ell \in R$, R is the group generating T , r_1 is the identity element in R and $|T|=J$.

III. AN EXAMPLE

We consider a case in which the message source has alphabet $M = \{1, 2, \dots, 7\}$ and T is a subgroup in G . The a priori probabilities of the message symbols are

$$\begin{aligned} q_1 &= 0,06 & q_2 &= 0,07 & q_3 &= 0,09 & q_4 &= 0,12 \\ q_5 &= 0,16 & q_6 &= 0,21 & q_7 &= 0,29. \end{aligned}$$

The entropy for the message source is $H(M) = 1.806$ Nats/Symb. T is defined in the following way: Let $v(\cdot)$ be an invertible mapping of M onto the nonzero 3-dimensional vectors with elements in $GF(2)$ and let $H = \{H_\ell\}_{\ell=1}^J$ be the set of all invertible 3×3 matrices over $GF(2)$. Then the elements in T are defined as

$$t_\ell(\cdot) = v^{-1} \left(v(\cdot) H_\ell \right) \quad \ell = 1, 2, \dots, J. \quad (4)$$

The specific choice of $v(\cdot)$ for this example, is the mapping for which the image of m can be interpreted as the binary representation of the number m , for example $v(3) = (0, 1, 1)$.

The number of elements in H is 168. Hence the number of keys is $J=168$.

In figure 2 we have plotted the upper bound of Theorem 1, the exact value of $H(K|E^L)$ (calculated for L even) and the lower bound given in Theorem 1 in [1], which in our case can be written as

$$H(K|E^L) \geq \log(168) - L[\log(7) - H(M)]. \quad (5)$$

IV. DISCUSSION

As is seen in the plot of the example, the upper bound has the same general behaviour as $H(K|\tilde{E}^L)$. Using the same technique as in [1] it is a straightforward exercise to show that when L goes to infinity both $H(K|\tilde{E}^L)$ and the upper bound has the same limit value. From (2) and (3) it is also clear that $H(K|\tilde{E}^L)$ and the upper bound have the same value for $L=0$.

The approach taken in [1] to show that the bound in Theorem 2, is exponentially tight does not work in general for this case. However, for certain pairs of sets of enciphering transformations and a priori probabilities of the message source the approach in [1] will work.

APPENDIX

To obtain (2) and (3) the steps and changes necessary in the derivations of the corresponding results in [1, eq. (22) and (27)] are summarized below. For easy crossreferencing, we number the equations in this Appendix with the number of the corresponding equation in [1].

The starting point in the derivation of (2) is

$$H(K|\tilde{E}^L) = \sum_{k=1}^J \sum_{\tilde{e}^L \in \tilde{E}^L} P_{E^L K}(\tilde{e}^L, k) \log \left(\frac{\sum_{\ell=1}^J P_{E^L K}(\tilde{e}^L, \ell)}{P_{E^L K}(\tilde{e}^L, k)} \right) \quad (14')$$

Because $t_k(\cdot)$ is invertible we have

$$P_{E^L K}(\tilde{e}^L, k) = \frac{1}{J} P_{M^L} (t_k^{-1}(\tilde{e}^L)) \quad (17')$$

The relation between R and T exhibited in (1) and substitution of $n=t_n^{-1}(t_\ell(n'))$ into eq. (18) in [1] gives

$$y_{t_\ell(n)} = x_{t_k^{-1}(t_\ell(n))} = x_{r_j(n)} \quad \text{for some } j, 1 \leq j \leq J \quad (18')$$

for the symbol frequencies in the cryptogram ($\underline{y}$) and the message ($\underline{x}$). Thus for a memoryless message source with symbol probabilities $\{q_n\}_1^N$ we have

$$P_{E^L K}(\tilde{e}^L, k) = \frac{1}{J} \prod_{n=1}^N q_n^{y_{t_k(n)}} = \frac{1}{J} \prod_{n=1}^N q_n^{x_n} \quad (19')$$

and

$$J \left(\sum_{\ell=1}^J P_{E^L K}(\tilde{e}^L, \ell) \right) = \sum_{\ell=1}^J \prod_{n=1}^N q_n^{y_{t_\ell(n)}} = \sum_{\ell=1}^J \prod_{n=1}^N q_{r_\ell}^{x_n} \quad (20')$$

Substitution of (19') and (20') into (14') gives (2).

As for (3), Lemma 2 in [1] applied to (14') gives

$$H(K|\tilde{E}^L) \leq \log \left(\sum_{e^L \in \tilde{E}^L} \sum_{k=1}^J \sum_{\ell=1}^J \sqrt{P_{E^L K}^L(e^L, k)} \sqrt{P_{E^L K}^L(e^L, \ell)} \right) \quad (29')$$

Rather the same substitutions as above give

$$H(K|\tilde{E}^L) \leq \log \left(\sum_{k=1}^J \frac{1}{J} \sum_{\ell=1}^J \left(\sum_{n=1}^N \sqrt{q_{t_k^{-1}(n)}^{-1}} q_{t_\ell^{-1}(n)}^{-1} \right)^L \right) \quad (30')$$

Substitution on $n' = t_\ell^{-1}(n)$ in (30') results in (cfr (18'))

$$\sum_{\ell=1}^J \left(\sum_{n=1}^N \sqrt{q_{t_k^{-1}(n)}^{-1}} q_{t_\ell^{-1}(n)}^{-1} \right)^L = \sum_{\ell=1}^J \left(\sum_{n'=1}^N \sqrt{q_{r_\ell(n')}^{-1}} q_{n'}^{-1} \right)^L \quad (31')$$

Then substitution of (31') in (30') gives (3) when we use the assumption that $r_1(\cdot)$ is the identity element in R .

Figure 1. Blockdiagram of the secrecy system

Figure 2. Plot of bounds on the equivocation of the key for the case considered in Section III

REFERENCES

- [1] R.J. Blom, "Bounds on Key Equivocation for Simple Substitution Ciphers", IEEE Trans. Inform. Theory, Vol. IT-25, No. 1, pp. 8-18, Jan. 1979.
- [2] R.J. Blom, "On Pure Ciphers", Internal Publication, LiTH-ISY-I-0286, Linköping University, 1979.