Problems of Control and Information Theory, Vol. 11 (4), pp. 301–318 (1982)

REMARKS ON SHANNON'S SECRECY SYSTEMS*

R. AHLSWEDE

(Bielefeld)

(Received August 20, 1981)

The paper contains three improvements of Shannon's theory of secrecy systems [1]:

1. By a very simple construction we obtain ciphers which are with respect to natural security measures as good as Shannon's "random ciphers".
2. For this construction it is unnecessary to assume that the messages are essentially equally likely. Shannon made this assumption in order to make his "random cipher" approach work.
3. Furthermore we construct optimal ciphers under the rather robust assumption that only a bound on the entropy of the source is known to the communicators and that the cryptanalyst is still granted to know the message statistic exactly.

Finally we construct worst codes for the binary symmetric channel and emphasize the importance of this "dual coding problem" for cryptography.

1. The model and an outline of the results

C. E. Channon presented in [1] a mathematical theory of secrecy systems. We now briefly describe his model and refer to the original paper for the reader interested in the heuristic behind the assumptions made.

We are given a source of messages $(\mathfrak{M}, P)$, where $\mathfrak{M} = \{1, \ldots, M\}$ and $P = (P_1, \ldots, P_M)$ is a probability distribution on $\mathfrak{M}$.

A key space $\mathcal{C} = \{c_1, \ldots, c_K\}$ is a set of bijective mappings (the keys) $c_i : \mathfrak{M} \to \{E_1, \ldots, E_M\}$, the set of cryptograms, which can be identified with $\mathfrak{M}$ via isomorphy. Thus the c_i's can be viewed as permutations on $\{1, \ldots, M\}$.

Let $Q = (Q_1, \ldots, Q_K)$ be a probability distribution on $\mathcal{C}$. The pair $(\mathcal{C}, Q)$ is called cipher.

The secrecy system operates as follows. $(\mathcal{C}, P)$ and $(\mathcal{C}, Q)$ are independent random experiments. First a key is selected according to $(\mathcal{C}, Q)$ and sent to the receiving point over a secure channel. Then a message is produced and the sender applies the key chosen to produce a cryptogram. This cryptogram

* Presented at the International Symposium on Information Theory, Ithaca, N. Y., Oct. 10–14, 1977.

is transmitted to the receiver by a noiseless channel and may be intercepted by the enemy cryptanalyst. The receiver can reproduce the message, because he knows the key, which is an invertible transformation. Since the cryptanalyst may eventually find out the message distribution anyhow, it is assumed right away that he knows it. For the same reason it is also assumed that he knows the cipher $(\mathcal{C}, Q)$. The identity of the key being used is the only information which is supposed to be unknown to him. The security of the system results from the fact that the cryptanalyst has several options for the key which might have been used. The main problem in this model is now to find subject to constraints, such as a limitation on the size K of the key space, ciphers which guarantee for a given source maximal security. To make this problem mathematically tractable a quantitative notion of security has to be introduced. Shannon used as criterion the average uncertainty about the message:

Let X, Y be random variables on $\mathfrak{M}$, resp. $\{E_1, \ldots, E_M\}$ (for simplicity we set $\{E_1, \ldots, E_M\} = \mathfrak{M}$ in the sequel) with joint distribution

$$\Pr(X = m,\ Y = m') = P_m \cdot \sum_{i : c_i(m) = m'} Q_i. \tag{1.1}$$

$\Pr(X = m) = P_m$ is the probability that message m is to be sent. The cryptanalyst intercepts the cryptogram $Y = m'$ with probability $\Pr(Y = m' | X = M)$, if m was sent. Observing Y the average uncertainty about X for him is the conditional entropy $H(X|Y)$. Measuring the "size" of cipher $(\mathcal{C}, Q)$ by $H(Q)$ Shannon suggested in [1] the study of the function

$$S(P, \alpha) = \max_{(C,Q) : H(Q) \leq \alpha} H(X|Y).$$

Another, but closely related measure of security was used by Hellman in [2]. He considers the average number of "spurious decipherments".

As a third secrecy measure one can consider the cryptanalyst's error probability in deciding upon the message sent. (During the presentation of this note we learnt that other fidelity criteria have been investigated in [5]). For given P, $\mathcal{C}$, Q the probability to decrypt correctly is given by

$$\lambda_c(P, \mathcal{C}, Q) = \sum_{m' \in M} \max_{m \in \mathfrak{M}} \Pr(Y = m', X = m),$$

where it is assumed again that the cryptanalyst uses the (best possible) maximum likelihood decision rule.

Considering this secrecy system, we address ourselves to the following questions.

1. Can good ciphers be constructed explicitly?
2. Is the hypothesis of essentially equidistributed messages (made in [1], page 691) necessary?

3. How can one deal with sources whose statistic is not known exactly? (Robustness).

In Section 2 we give as an answer to question 1 a very simple construction of a cipher which is with respect to any security measure as good as Shannon's "random ciphers". Moreover, the construction gives also an answer to question 2: the hypothesis of essentially equidistributed messages is completely unnecessary. It was needed in [1] to make the "random cipher" approach work. Since for many stochastic processes the AEP-property does not hold, one should try to avoid this hypothesis.

The construction in Section 2 leads to almost optimal, however, not strictly optimal ciphers for all distributions $P = (P_1, \ldots, P_M)$ on the messages. In Section 3 we make an attempt to construct best ciphers with respect to the error probability criterion. The only results of some general nature is Lemma 1. It helps for the construction of those ciphers. The approach gives satisfactory insight only in the case of two keys.

With respect to question 3 we suggest to make Shannon's model more robust in the following sense: We completely drop the assumption that the message statistic is known to the communicators and still grant the cryptanalyst that he knows it. The performance of a cipher depends decisively on the entropy $H(X)$ of the source. If for instance the entropy is 0 and therefore all probability concentrated on one message, the cryptanalyst just votes for that message and no secrecy is possible, no matter how the cipher is chosen. Using "random ciphers" Shannon derived for a known source $(\mathfrak{M}, P)$ with the additional hypothesis of equidistribution on a large probability subset of messages the bound

$$H(X \mid Y) \geq \log K + H(X) - \log M.$$

In Section 2 we see that this bound is very poor for a general X and good only if $H(X) = \log M$, that is, the source is compressed. The reason for this poor performance of random ciphering is due to the fact that keys are chosen at random from *all* permutations on $\mathfrak{M}$ and therefore small probability messages are exchanged with high probability messages. In our robust model in Section 4 we permit all source statistics X with entropy $H(X) \geq H_0$. We show that with one cipher one can achieve

$$H(X \mid Y) \geq \log K + H_0 - \log M$$

for all X with $H(X) \geq H_0$.

Moreover, this bound is tight for the class and thus finds its natural place.

Finally, as a first contribution to the dual coding problem we construct worst codes for the binary symmetric channel in Section 5. It was already

emphasized in [2] that the problem of ciphering is in a sense dual to the problem of coding. Here the "code" is given and we look for channels subject to constraints for which this code is worst. There is, however, a noticeable difference between the two problems: whereas good codes are hard to find, worst codes can be constructed explicitly for the binary symmetric channel (BSC) as we shall show in Section 5. It therefore seems to be unnecessary to escape to random choices of secrecy systems unlike the situation in coding theory where random coding is often the only method to quarantee at least the existence of codes with certain parameters.

2. A simple cipher for Shannon's secrecy system

Let $(\mathfrak{M}, P)$ be a message source, where $\mathfrak{M} = \{1, \ldots, M\}$ and $P = (P_1, \ldots, P_M)$ is probability distribution on $\mathfrak{M}$.

We shall present a cipher with K keys $c_1, \ldots, c_K$, which is very good with respect to both security measures defined in Section 1. If $K \geq M$ one can choose the M keys $c_i(m) = m + i \bmod M$ $(i = 1, \ldots, M)$ with equal probability and then for this cipher $H(X) = H(X \mid Y)$, where X, Y are defined as in (1.1). Therefore we can always assume $K \leq M$. We also consider here only *canonical* ciphers, that is, ciphers whose keys are equiprobable. In [1] (page 691), and in [2] (page 6) ciphers were also assumed to be canonical. This restriction does not seem to be severe, but this has to be proved. We call a cipher $(\mathcal{C}, Q)$ *regular* if for all $m \in \mathfrak{M}$:

$$|\{c_i(m) | 1 \leq i \leq K\}| = |\{c_i^{-1}(m) | 1 \leq i \leq K\}| = K. \tag{2.1}$$

We define now a simple regular cipher for the given $(\mathfrak{M}, P)$, where we assume without loss of generality that

$$P_1 \geq P_2 \geq P_3 \geq \ldots \geq P_M. \tag{2.2}$$

We write M in the form

$$M = K \cdot l + r, \; 0 \leq r < K \tag{2.3}$$

and we define K keys $c_0, \ldots, c_{K-1}$ as follows

$$c_i(m) = \begin{cases} Kj + ((t + i - 1) \bmod K) + 1 \text{ if } m = Kj + t \text{ with} \\ \qquad 0 \leq j \leq l - 2, \; 1 \leq t \leq K \\ K(l - 1) + ((t + i - 1) \bmod K + r) + 1 \text{ if} \\ \qquad m = K(l - 1) + t \text{ with } 1 \leq t \leq K + r. \end{cases} \tag{2.4}$$

These transformations map the blocks of messages

$$B_1 = \{1, \ldots, K\}, \quad B_2 = \{K + 1, \ldots, 2K\}, \ldots,$$

$$B_{l-1} = \{K(l-2)+1, \ldots, K(l-1)\}, \quad B_l = \{K(l-1)+1, \ldots, Kl+r\}$$

onto themselves. Obviously (2.1) holds and the cipher is regular.

Let $\mathcal{C} = \{c_0, \ldots, c_{K-1}\}$ and let Q be the equidistribution on $\mathcal{C}$. Then $(\mathcal{C}, Q)$ is a regular canonical cipher.

If the cryptanalyst intercepts a cryptogram $m' \in \mathfrak{M}$, then with respect to the error probability criterion the best decoding rule (maximum likelihood) is to vote for

$$Kj+1, \text{ if } m' \in B_{j+1} \text{ and } 0 \leq j \leq l-2$$

$$Kj+1, \text{ if } m' \in \{K(l-1)+1, \ldots, Kl\}, \text{ resp. for}$$

$$K(l-1)+t, \text{ if } m' = lK+t-1 \text{ and } 2 \leq t \leq r+1.$$

Then the messages $\{Kj+1 \mid 0 \leq j \leq l-1\}$ are always decrypted correctly, the messages $\{K(l-1)+t \mid 2 \leq t \leq r+1\}$ are decrypted correctly with probability $\frac{1}{K}$, and all other messages are always decrypted incorrectly. Therefore the error probability λ for this cipher can be expressed as

$$\lambda = 1 - P_1 - P_{K+1} - \ldots - P_{K(l-1)+1} - \\ - \frac{1}{K} P_{K(l-1)+2} - \ldots - \frac{1}{K} P_{K(l-1)+r+1}. \tag{2.5}$$

Therefore

$$\lambda \geq (P_2 + \ldots + P_K) + (P_{K+2} + \ldots + P_{2K}) + \ldots + (P_{K(l-2)+2} + \ldots + \\ + P_{K(l-1)}) + \frac{K-1}{K}(P_{K(l-1)+2} + \ldots + P_{K(l-l)+r+1}) + (P_{K(l-1)+r+2} + \ldots P_{Kl+r}).$$

This and (2.2) imply

$$\lambda \geq (K-1)\left(P_K + P_{2K} + \ldots + P_{K(l-1)} + \\ + \frac{1}{K} P_{K(l-1)+2} + \ldots + \frac{1}{K} P_{K(l-1)+r+1}\right) \tag{2.6}$$

and (2.6), (2.5), and again (2.2) yield

$$\lambda \geq (K-1)(1 - \lambda - P_1)$$

and thus

$$\lambda \geq \frac{K-1}{K}(1 - P_1).$$

Since the cryptanalyst can always vote for message $1 \in \mathfrak{M}$, obviously also

$$\lambda \leq 1 - P_1$$

must hold. This shows that for large K, $\lambda \geq \frac{K-1}{K}(1 - P_1)$ is close to the "optimum" $1 - P_1$.

In summary we have

Theorem 1. For the cipher described in (2.4) the decrypting error probability λ satisfies

$$\frac{K-1}{K}(1 - P_{\max}) \leq \lambda \leq 1 - P_{\max},$$

where $P_{\max} = \max_{m \in \mathfrak{M}} P_m$.

We consider now the entropy criterion $H(X \mid Y)$.

In [1] Shannon assumed that the set of possible messages $\mathfrak{M}$ can be divided into two groups: one group of high and fairly uniform probability, the second of negligibly small total probability. If the high probability group has cardinality N, then we have $H(X) \sim \log N$. Using a cipher selected at random Shannon gave for this situation the bound

$$H(X \mid Y) \geq \log K + H(X) - \log M. \tag{2.7}$$

Note that (2.7) trivially holds for *every* regular cipher, without any assumption on X. It is amazing that a better result can be obtained by evaluating $H(X \mid Y)$ for our simple cipher.

We derive first a bound by standard manipulations of entropy quantities, which is already better than (2.7). Then we show that for (also infinite) discrete probability distributions satisfying the rather natural assumption

$$P_i \leq \frac{1}{K} \qquad (i = 1, 2, \ldots). \tag{2.8}$$

$H(X \mid Y)$ is very close to $\log K$.

Notice that in addition to (2.8) no equidistribution property will be assumed and the result is essentially best possible, because always $H(X \mid Y) \leq \leq \log K$. Also, the hypothesis (2.8) cannot be weakened significantly. If $P_{i_0} = 1/2$ for some i_0 for instance, the cryptanalyst can always vote for i_0 and thus keep $H(X \mid Y)$ significantly smaller than $\log K$.

In order to give the first bound we define a random variable Z with l values and distribution

$$\Pr(Z = j) = \sum_{t=1}^{K} P_{K(j-1)+t} \quad \text{for } j = 1, \ldots, l-1, \text{ and}$$

$$\Pr(Z = l) = \sum_{t=1}^{K+r} P_{K(l-1)+t}.$$

Further we define random variables X_j, $j = 1, \ldots, l-1$ with distributions

$$\Pr(X_j = K(j-1) + t) = \frac{P_{K(j-1)+t}}{\Pr(Z = j)}, \qquad 1 \leq t \leq K.$$

Finally we define a random variable X_l with distribution

$$\Pr(X_l = K(l-1) + t) = \frac{P_{K(l-1)+t}}{\Pr(Z = l)}, \qquad 1 \leq t \leq K + r.$$

We shall prove

Theorem 2'. For the cipher described in (2.4) the cryptanalyst's uncertainty $H(X \mid Y)$ satisfies

$$H(X \mid Y) \geq \sum_{j=1}^{l} \Pr(Z = j) H(X_j) - \Pr(Z = l) \log \frac{K + r}{K}. \tag{2.10}$$

Proof. We use $H(Y \mid X) = \log K$ (regularity), $H(X \mid Y) + H(Y) = H(X) + H(Y \mid X)$, and the grouping axiom of the entropy function to get

$$H(X \mid Y) + H(Y) = H(Z) + \sum_{j=1}^{l} \Pr(Z = j) H(X_j) + \log K. \tag{2.11}$$

Since Y is equidistributed on the blocks $B_1, \ldots, B_{l-1}$

$$H(Y) \leq H(Z) + \sum_{j=1}^{l} \Pr(Z = j) \log K + \Pr(Z = l) \log (K + r).$$

This and (2.11) imply (2.9), which in turn yields (2.10).

Now we permit $\mathfrak{M}$ to be infinite. Then there is no last block, that is, $l = \infty$. Otherwise our cipher is defined as before and (2.9) takes the form

$$H(X \mid Y) = \sum_{j=1}^{\infty} \Pr(Z = j) H(X_j) = H(X \mid Z). \tag{2.12}$$

Next we prove

Theorem 2.[2] Let K be the number of keys.

[2] Originally we derived the bound $\log K - 7$. Simplifying our proof, I. Csiszár obtained the present bound.

Suppose that $(P_1, P_2, \ldots)$ is a discrete message distribution with

$$\frac{1}{K} \geq P_i \quad \text{for } i \in \{1, 2, \ldots\}.$$

then for our simple cipher

$$H(X \mid Y) \geq \log K - 1.$$

Proof. By (2.12) it suffices to give a lower bound on $H(X \mid Z)$. For this we write $\Pr(Z = j)$ in the form

$$\Pr(Z = j) = \frac{1}{K^{\varepsilon_j}}, \quad \text{where } 0 \leq \varepsilon_1 \leq \varepsilon_2 \leq \ldots. \tag{2.13}$$

Let us look at the first block. Since its total probability equals $\Pr(Z = 1) = \frac{1}{K^{\varepsilon_1}}$ and since the individual probabilities are smaller than $\frac{1}{K}$ by the monotonicity properties of $x \log x$

$$H(X \mid Z = 1) \Pr(Z = 1) \geq \frac{1}{K^{\varepsilon_1}} \log K^{1-\varepsilon_1}.$$

By the monotonicity of the P_i's

$$P_{K+1} \leq \frac{1}{K^{\varepsilon_1}} K^{-1} = \frac{1}{K^{1+\varepsilon_1}},$$

and repetition of the previous argument gives

$$H(X \mid Z = 2) \cdot \Pr(Z = 2) \geq \frac{1}{K^{\varepsilon_2}} ((1 + \varepsilon_1 + \varepsilon_2) \log K - 1).$$

By reiteration therefore

$$H(X \mid Y) \geq \sum_{j=1}^{\infty} \frac{1}{K^{\varepsilon_j}} ((1 + \varepsilon_{j-1} - \varepsilon_j)) \log K \tag{2.14}$$

(with the convention $\varepsilon_0 = 0$).

Of course also

$$\sum_{j=1}^{\infty} \frac{1}{K^{\varepsilon_j}} = 1. \tag{2.15}$$

These two relations imply

$$H(X \mid Z) \geq \log K - \sum_{j=1}^{\infty} \frac{\varepsilon_j - \varepsilon_{j-1}}{K^{\varepsilon}} \log K. \tag{2.16}$$

Since for natural logarithms $\log x \leq x - 1$, we have that

$$(\varepsilon_j - \varepsilon_{i-1}) \log K \leq K^{\varepsilon_j - \varepsilon_{j-1}} - 1,$$

which is equivalent to

$$\frac{\varepsilon_j - \varepsilon_{j-1}}{K^{\varepsilon_j}} \leq (\log K)^{-1} \left(\frac{1}{K^{\varepsilon_{j-1}}} - \frac{1}{K^{\varepsilon_j}} \right).$$

We can conclude that

$$\sum_{j=1}^{\infty} \frac{\varepsilon_j - \varepsilon_{j-1}}{K^{\varepsilon_j}} \leq \frac{1}{\log K},$$

and the Theorem is proved.

3. On the construction of optimal ciphers

The construction of best ciphers with réspect to the probability of error criterion furnishes an interesting combinatorial extremal problem. Our only result of a general nature is Lemma 1 below. It is shown for the case $K = 2$ and for canonical ciphers how this lemma can be used for the construction of ciphers.

For $m, m' \in \mathfrak{M}$ and a cipher $(\mathcal{C}, Q)$ we say that m and m' are connected and write (m, m'), if there exists a $c \in \mathcal{C}$ with $m' = c(m)$. $\overline{(m, m')}$ will mean that m and m' are not connected.

We mention again that for regular canonical ciphers a best decrypting rule for the cryptanalyst is, after having intercepted m', to vote for a message m of largest probability among those connected with m'.

Let $(\mathfrak{M}, P)$ with $P_1 \leq P_2 \leq \ldots \leq P_M$ and a cipher $(\mathcal{C}, Q)$ be given. Since we are concerned only with canonical ciphers in this section we write simply $\mathcal{C}$ instead of $(\mathcal{C}, Q)$. We denote by D_m $(m = 1, \ldots, M)$ the set of elements decrypted into m.

For $m \in \{1, \ldots, M\}$ we define

$$I_{\mathcal{C}}(m) = \{m' \in \mathfrak{M} \mid m' \in D_M, (m, m')\}$$

and

$$0_{\mathcal{C}}(m) = \{m' \in \mathfrak{M} \mid m' \notin D_M, (m, m')\}.$$

Denote by $m(\mathcal{C})$ the smallest $m \in \mathfrak{M}$ with $I_{\mathcal{C}}(m) \neq \emptyset$.

Lemma 1. There exists a cipher $\mathcal{C}$, which is optimal among the regular canonical ciphers, such that

a) $$|I_{\mathcal{C}}(m(\mathcal{C}))| \leq |I_{\mathcal{C}}(m(\mathcal{C})) + 1| \leq \cdots \leq |I_{\mathcal{C}}(M)|,$$

and

b) for any $\tilde{m}, m$; $\tilde{m} > m$; either $I_{\mathcal{C}}(\tilde{m}) \supset I_{\mathcal{C}}(m)$ or $\emptyset \neq 0_{\mathcal{C}}(\tilde{m}) \subset 0_{\mathcal{C}}(m)$.

Proof. Denote by $\mathfrak{R}$ the set of all optimal regular canonical ciphers on $\{1, \ldots, M\}$.

Let $m_1 = \max_{\mathcal{C} \in \mathfrak{R}} m(C)$ and define

$$\mathfrak{R}(m_1) = \{\mathcal{C} \in \mathfrak{R} \mid m(\mathcal{C}) = m_1, |I_{\mathcal{C}}(m_1)| = \min_{\mathcal{C}' : m(\mathcal{C}') = m_1} |I_{\mathcal{C}'}(m_1)|\}.$$

Let $m_2 = m_1 + 1$, $m_3 = m_2 + 1, \ldots, m_{M-1} = m_{M-2} + 1 = m_1 + M - 1$.

We define

$$\mathfrak{R}(m_1, m_2) = \left\{\mathcal{C} \in \mathfrak{R}(m_1) \,||I_{\mathcal{C}}(m_2)| = \min_{\mathcal{C}' \in \mathfrak{R}(m_1)} |I_{\mathcal{C}'}(m_2)|\right\}$$

etc. ... until

$$\mathfrak{R}(m_1, m_2, \ldots, M) = \left\{\mathcal{C} \in \mathfrak{R}(m_1, \ldots, M-1) \,|\, I_{\mathcal{C}}(M) = \min_{\mathcal{C}' \in \mathfrak{R}(m_1 \cdots, M-1)} |I_{\mathcal{C}'}(M)|\right\}.$$

Obviously, $\mathfrak{R}(m_1, \ldots, M) \neq \emptyset$.

We choose a $\mathcal{C}_1 \in \mathfrak{R}(m_1, \ldots, M)$. We show that the conditions a) and b) of the Lemma hold for $\mathcal{C}_1$.

a) Assume that for $m_1 \leq m \leq \tilde{m}$:

$$|I_{\mathcal{C}_1}(m)| > |I_{\mathcal{C}_1}(\tilde{m})|.$$

Then there exists an $m' \notin D_M$ and an $m'' \in D_M$ such that

$$(\tilde{m}, m'), (m, m''), \overline{(m, m')}, \text{ and } \overline{(\tilde{m}, m'')} \text{ hold.}$$

Replace $(\tilde{m}, m')$, (m, m'') by (m, m'), $(\tilde{m}, m'')$. Then we get a cipher $\mathcal{C}_2$ with

$$|I_{\mathcal{C}_2}(m)| = |I_{\mathcal{C}_1}(m)| - 1, \tag{3.1}$$

and

$$|I_{\mathcal{C}_2}(\tilde{m})| = |I_{\mathcal{C}_1}(\tilde{m})| \text{ for any } \tilde{m} < m. \tag{3.2}$$

Since $P_{m_1} \leq P_m \leq P_M$ it is clear from the definition of the error probability that the error probability for the new cipher is not smaller than the original one. This and (3.1), (3.2) contradict the definition of $\mathcal{C}_1$.

b) If $I_{\mathcal{C}_1}(m) = \emptyset$ the statement is obviously true. If condition b) does not hold for $\mathcal{C}_1$, then there exist m', m'' such that

$$m' \notin I_{\mathcal{C}_1}(\tilde{m}),\ m' \in I_{\mathcal{C}_1}(m),\ m'' \in 0_{\mathcal{C}_1}(\tilde{m}),\ m'' \notin 0_{\mathcal{C}_1}(m).$$

Replace $(\tilde{m}, m'')$, (m, m') by (m, m''), $(\tilde{m}, m')$. Again the decoding error probability does not decrease because m' is decrypted into M anyhow and m'' has now besides the old connections a connection with an element of smaller probability.

We use now this Lemma to obtain best regular canonical ciphers for the case $K = 2$.

For canonical ciphers $\mathcal{C} = \{c_1, \ldots, c_k\}$ the incidence matrix $\mathfrak{J}$ is an $M \times M$ matrix with entries from $\{0, 1\}$ such that the i, j-entry of this matrix is 1 iff i and j are connected by the cipher, that is iff (i, j).

The incidence matrix $\mathfrak{J}$ of a cipher $\mathcal{C}$ is a unique representation of $\mathcal{C}$. For the special case $K = 2$ we look now for best incidence matrices.

For $K = 2$ any regular cipher $\mathcal{C}$ satisfies of course $|D_M| = 2$ and there are exactly four connections with the two elements of D_M. Thus we can conclude from Lemma 1 (a) that either (w. l. o. g. $D_M = \{M - 1, M\}$)

$$|I_{\mathcal{C}}(M - 1)| = |I_{\mathcal{C}}(M)| = 2 \qquad \text{(Case 1)}$$

or

$$|I_{\mathcal{C}}(M - 2)| = |I_{\mathcal{C}}(M - 1)| = 1,\ |I_{\mathcal{C}}(M)| = 2. \qquad \text{(Case 2)}$$

We make the following convention: For matrices $\mathfrak{J}_1$, $\mathfrak{J}_2$ we denote by $[\mathfrak{J}_1, \mathfrak{J}_2]$ the matrix

$$[\mathfrak{J}_1, \mathfrak{J}_2] = \begin{pmatrix} \mathfrak{J}_1 & 0 \\ 0 & \mathfrak{J}_2 \end{pmatrix},$$

where 0 is a zero-matrix.

Case 1. In this case obviously the incidence matrix $\mathfrak{J}$ of $\mathcal{C}$ has the structure

$$\mathfrak{J} = [\tilde{\mathfrak{J}}_1, \mathfrak{J}_1]$$

with $\mathfrak{J}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

Case 2. Since $K = 2$ and M picks up two connections with D_M, necessarily

$$I_{\mathcal{C}}(M - 1) \cap I_{\mathcal{C}}(M - 2) = \emptyset.$$

By Lemma 1b) therefore $O_{\mathcal{C}}(M-1) \subset O_{\mathcal{C}}(M-2)$ and hence

$$O_{\mathcal{C}}(M-1) = O_{\mathcal{C}}(M-2),$$

because $|O_{\mathcal{C}}(M-1)| = |O_{\mathcal{C}}(M-2)|$. We conclude in this case: $\mathfrak{J}$ has the structure (w. l. o. g. $I_l(M-1) = \{M\}$ and $O_l(M-1) = \{M-2\}$)

$$\mathfrak{J} = [\tilde{\mathfrak{J}}, \mathfrak{J}_2]$$

with $\mathfrak{J}_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

By iteratively applying the above argument we get the result:

Theorem 3. For $K = 2$ and arbitrary $M \geq 2$ there exists an optimal regular canonical cipher of the form

$$[\mathfrak{J}_{\varepsilon_1}, \mathfrak{J}_{\varepsilon_2}, \ldots, \mathfrak{J}_{\varepsilon_t}], \text{ where } \varepsilon_1, \ldots, \varepsilon_t \in \{1, 2\}.$$

For small values of M there are the following possibilities:

$$M = 2 : \mathfrak{J} = \mathfrak{J}_1$$

$$M = 3 : \mathfrak{J} = \mathfrak{J}_2$$

$$M = 4 : \mathfrak{J} = [\mathfrak{J}_1, \mathfrak{J}_1]$$

$$M = 5 : \mathfrak{J} = [\mathfrak{J}_1, \mathfrak{J}_2] \text{ or } [\mathfrak{J}_2, \mathfrak{J}_1]$$

$$M = 6 : \mathfrak{J} = [\mathfrak{J}_1, \mathfrak{J}_1, \mathfrak{J}_1] \text{ or } [\mathfrak{J}_2, \mathfrak{J}_2]$$

$$M = 7 : \mathfrak{J} = [\mathfrak{J}_2, \mathfrak{J}_1, \mathfrak{J}_1] \text{ or } [\mathfrak{J}_1, \mathfrak{J}_2, \mathfrak{J}_1] \text{ or } [\mathfrak{J}_1, \mathfrak{J}_1, \mathfrak{J}_2].$$

By a suitable choice of the source probabilities $P = (P_1, \ldots, P_M)$ one can show that all these cases occur as best incidence matrices. We show now that for $K = 2$ nothing is gained if we drop the supposition of regularity. This can be readily verified as follows:

We can assume that $|D_M| = 2$, because otherwise M is doubly connected with one element m' say, and the cipher can be improved by replacing any $(\tilde{m}, m'')$, $\tilde{m}, m'' \neq M$ and one connection (M, m') by $(\tilde{m}, m')$ and (M, m'').

If there is an m with $|I_{\mathcal{C}}(m)| = 2$ we are in Case 1: $\mathfrak{J}_1$. Otherwise we have two elements $m_1, m_2 \neq M$, $P_{m_2} \geq P_{m_1}$, with $|I_{\mathcal{C}}(m_1)| = |I_{\mathcal{C}}(m_2)| = 1$ and $I_{\mathcal{C}}(m_1) \cap I_{\mathcal{C}}(m_2) = \emptyset$. Then either we are in Case 2: $\mathfrak{J}_2$ or we have for $m', m'' \notin D_M$ (m_1, m') and (m_2, m''). Since there is also an $m^* \in D_M$ with (m_1, m^*) we can

replace (m_1, m^*) by (m_2, m^*) and (m_2, m'') by (m_1, m''). The error probability does not decrease. We are again in Case 1: $\mathfrak{J}_1$.

Remarks. These considerations might suggest that for $K > 2$ there is a fixed number $f(K)$ of incidence matrices, independent of $(\mathfrak{M}, P)$ such that optimal incidence matrices can be built from these few basic matrices. Unfortunately, already for $K = 3$ this conjecture is false. Also for $K = 3$ there are non-regular ciphers which are better than best regular ciphers.

However, a result similar to Theorem 6 can be obtained *for arbitrary* K if one imposes a kind of convexity condition on the probability vector $P = (P_1, \ldots, P_M)$. More precisely, if $P_1 \leq \ldots \leq P_M$ and

$$P_m \leq \frac{P_{m-1} + P_{m+1}}{2}$$

for any $m \in \{2, \ldots, M-1\}$, then there are two incidence matrices $\mathfrak{J}_1^K, \mathfrak{J}_2^K$ such that a matrix of the form

$$[\mathfrak{J}_1^K, \mathfrak{J}_1^K, \mathfrak{J}_1^K, \ldots, \mathfrak{J}_1^K, \mathfrak{J}_2^K]$$

represents a best reguar canonical cipher. This was proved by Blome and Jusek in their (unpublished) diplom thesis.

4. A robustification of Shannon's secrecy system

From the definition of regularity of a cipher one can see immediately that every regular canonical cipher satisfies

$$\begin{aligned} H(X \mid Y) &= H(Y \mid X) + H(X) - H(Y) \\ &= \log K + H(X) - H(Y) \\ &\geq \log K + H(X) - \log M \end{aligned}$$

for all message variables X, and therefore also

$$H(X \mid Y) \geq \log K + H_0 - \log M$$

for all message variables X with $H(X) \geq H_0$. We show now that this bound is essentially best possible for canonical ciphers:

Theorem 4. For every canonical cipher $(\mathcal{C}, Q)$ on $\mathfrak{M} = \{1, \ldots, M\}$ with K keys and for every H_0, $0 \leq H_0 \leq \log M$, there exists a message variable X with values in $\mathfrak{M}$, $H(X) \geq H_0$, such that

$$(4.1)\quad H(X \mid Y) \geq [\log K + H_0 - \log M]^+ + \log \frac{4}{\varepsilon} + h(\varepsilon) + \varepsilon \log K,$$

where $0 < \varepsilon < \frac{1}{2}$ and $[t]^+ = \max\{t, 0\}$.

For the proof of Theorem 4 we need an auxiliary result, which is readily obtained by using Feinstein's maximal coding idea for the construction of a code with code words from a prescribed subset $\mathcal{A} \subset \mathfrak{M}$.

Recall that for a message source $(\mathfrak{M}, P)$ and a cipher $(\mathcal{C}, Q)$ we defined random variables X, Y in (1.1). Now we consider

$$W(m' \mid m) = \Pr(Y = m' \mid X = m);\ m', m \in \mathfrak{M},$$

as a transmission matrix of a "channel" associated with $(\mathfrak{M}, P)$ and $(\mathcal{C}, Q)$. We use again the notion of connectedness of two messages, which was introduced in Section 3.

Lemma 2. Let W be the transmission matrix associated with a canonical cipher on $\mathfrak{M} = \{1, \ldots, M\}$ and let $\mathcal{A} \subset \mathfrak{M}$ be such that $|\mathcal{A}| \geq (1 - \delta) M$, $0 < \delta < 1$. Then for any ε, $0 < \varepsilon < \frac{1}{2}$ there exists an ε-code $\{(u_i, D_i) \mid i = 1, \ldots, N\}$ for W such that $\{u_i \mid i = 1, \ldots, N\} \subset \mathcal{A}$ and $N = \left[\frac{\varepsilon}{K}(1 - \delta) M\right]$. $[t]$ is the integral part of t.

Proof. Let $\{(u_i, D_i) \mid i = ,\ldots, N\}$ be an ε-code with $\{u_i \mid i = 1, \ldots, N\} \subset \mathcal{A}$, such that u_i is connected with every element of D_i and such that the code cannot be prolonged in $\mathcal{A}$. Then for all $u \in \mathcal{A}$ $W\left(\bigcup_{i=1}^{N} D_i \mid u\right) > \varepsilon$ and therefore $\left|\bigcup_{i=1}^{N} D_i\right| K \geq \varepsilon\, |\mathcal{A}| \cdot K$. Hence, $N \cdot K \geq \left|\bigcup_{i=1}^{N} D_i\right| \geq \varepsilon\, |\mathcal{A}|$ and $N \geq \frac{\varepsilon}{K}(1 - \delta) M$.

Proof of Theorem 4. By iteratively applying Lemma 2 we can construct ε-codes $\{(u_i^t, D_i^t) \mid i = 1, \ldots, N\}$ for $t = 1, \ldots, T$ with all u_i^t distinct provided that

$$T \cdot \frac{\varepsilon}{K}(1 - \delta) M \leq \delta M.$$

This is satisfied, if $T \leq \frac{\delta}{\varepsilon(1 - \delta)} K$. Define now a random variable X with distribution

$$Pr(X = u_j^t) = \frac{1}{NT}.$$

Let Y be the corresponding output variable with respect to W. By the grouping axiom for the entropy function and by Fano's Lemma we get

$$H(X \mid Y) \leq \log T + h(\varepsilon) + \varepsilon \log K. \tag{4.2}$$

Actually, Fano's Lemma applied directly would give only a term $\varepsilon \log N$, here we can do better because every $m' \in \mathfrak{M}$ is connected with at most K u's in a code.

Now we choose T as small as possible under the condition $\log (T \cdot N) \geq H_0$

Clearly, $\log T \leq H_0 - \log \frac{\varepsilon}{K} - \log (1 - \delta) - \log M + 1$ and (4.2) yields for $\delta = \frac{1}{2}$

$$H(X \mid Y) \leq H_0 + \log K - \log M + \log \frac{2}{\varepsilon} + h(\varepsilon) + \varepsilon \log K + 1$$

which is (4.1).

5. Worst codes for the BSC

The result presented below is for binary symmetric channels with transmission matrix $W = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$, $0 \leq \varepsilon \leq \frac{1}{2}$, its extension to general DMC's seems to be an interesting mathematical problem.

Coding theory has been concerned with the problem to find (n, R)-codes, i.e., codes of block length n and rate R, for which the average error probability is small. Nobody found for arbitrary n and positive rate codes which are optimal in the sense that error probability assumes its minimum. This is a very hard combinatorial extremal problem and has led to numerous investigations in probabilistic and algebraic coding theory.

We study here the dual problem: find (n, R) codes with *distinct* code words for which the decoding error probability is maximal. More generally we also permit an arbitrary message statistic rather than just the equidistribution.

The problem then takes the following form: Given a probability distribution $P = (P_1, \ldots, P_{2^n})$ on 2^n elements, find a bijective map $U : \{1, \ldots, 2^n\} \to \{0, 1\}^n$ such that

$$\lambda_c(P) = \max_{\mathfrak{D}} \sum_{i=1}^{2n} P_i \cdot W^n(D_i \mid u_i) \tag{5.1}$$

is minimal. Here $u_i = U(i)$; $W^n(\cdot|\cdot)$ denotes the n-fold product of the transmission probability function of the BSC, and $\mathfrak{D} = \{D_1, \ldots, D_{2^n}\}$ is a decoding rule.

We describe now an explicit solution to the problem. W.l.o.g. we can assume that

$$P_1 \geq P_2 \geq \ldots \geq P_{2^n}.$$

Let us order the vectors v in $\{0, 1\}^n$ primarily according to the number of components with value 0 and secondarily lexicographically, where 1 precedes 0. Thus

$$v_1 \geq v_2 \geq \ldots \geq v_{n+1} \geq v_{n+2} \geq \ldots \geq v_{\binom{n}{2}+n+1} \geq \ldots \geq v_{2^n}.$$

Theorem 5. Let $P = (P_1, \ldots, P_{2^n})$ be a probability distribution on the messages, $P_i \geq P_{i+1}$, then the encoding $U(i) = v_i$ for $i = 1, \ldots, 2^n$ minimizes the probability of correct decoding $\lambda_c(P)$ (as defined in (5.1)).

For (n, R) codes one gets the solution to th above problem by choosing $P_i = \dfrac{1}{N}$ for $i = 1, \ldots, N = [e^{nR}]$.

For the proof of Theorem 4 we need an extension of a result of Harper [3]. Let us denote by $S_r(x^n)$ the Hamming sphere in $\{0, 1\}^n$ with center $x^n \in \{0, 1\}^n$ and radius r.

Then we have:

General isoperimetry theorem. Let $\{r_i\}_{i=1}^N$ be a decreasing sequence of integers, then for any distinct $x_1^n, \ldots, x_N^n \in \{0, 1\}^n$:

$$\left|\bigcup_{i=1}^N S_{r_i}(x_i^n)\right| \geq \left|\bigcup_{i=1}^N S_{r_i}(v_i)\right|.$$

Harper [3] proved this in the case $r_i = r$, $i = 1, \ldots, N$. We show here that the given general case easily follows from his result.

Proof. Fix any $j \in \{0, \ldots, N-1\}$. Then for any $i \in \{1, \ldots, N-j\}$ $i \leq N - j$ holds and by the monotonicity of the radii we have for those i

$$r_i \geq r_{N-j} \text{ and } |S_{r_i}(x_i^n)| \geq |S_{r_{N-j}}(x_i^n)|.$$

Hence,

$$\left|\bigcup_{i=1}^N S_{r_i}(x_i^n)\right| \geq \max_{j \in \{0, \ldots, N-1\}} \left|\bigcup_{i=1}^{N-j} S_{r_{N-j}}(x_i^n)\right|.$$

By Harper's theorem the expression on the right-hand side is minimal if $x_i^n = c_i$ for $i = 1, \ldots, N$. Furthermore it can easily be verified that $\bigcup_{i=1}^{N-j} S_{r_{N-j}}(v_i)$ equals $\{v, \ldots, v_{t_j}\}$ for a suitable t_j.

Therefore, there is a $j' \in \{0, \ldots, N-1\}$ such that $\bigcup_{i=1}^{N-j'} S_{r_{N-j'}}(v_i)$ contains all the sets $\bigcup_{i=2}^{N-j} S_{r_{N-j}}(v_i)$, $j \in \{0, \ldots, N-1\}$. We conclude that

$$\left| \bigcup_{i=1}^{N} S_{r_i}(v_i) \right| = \left| \bigcup_{j=0}^{N-j} \left(\bigcup_{i=1}^{N-j} S_{r_{N-j}}(v_i) \right) \right| = \max_{j \in \{0, \ldots, N-1\}} \left| \bigcup_{i=1}^{N-j} S_{r_{N-j}}(v_i) \right|$$

which proves the theorem.

Proof of Theorem 5. For a map $U : \{1, \ldots, 2^n\} \to \{0, 1\}^n$ $(U(i) := u_i)$, a decoding rule is optimal iff

$$D_i \subset \{y^n \in \{0, 1\}^n \mid P_i \beta^{d(y^n, u_i)} \geq P_j \beta^{d(y^n, u_j)} \text{ for all } j\} \tag{5.2}$$

and $\bigcup_{i=1}^{2^n} D_i = \{0, 1\}^n$, where $\beta = \dfrac{\varepsilon}{1-\varepsilon} \leq 1$ and where $d(\cdot)$ denotes the Hamming distance. Note that in (5.2) we have formulated just the concept of maximum likelihood decoding for the special case of the BSC. It should be clear intuitively that best decoding sets for the code word u_i are "like spheres around u_i", the diameter of which depends on P_i. We make this heuristic precise and apply the general isoperimetry theorem. For $y^n \in \{0, 1\}^n$ define

$$m(y^n, U) = \max_i P_i \cdot \beta^{d(y^n, u_i)}.$$

Then our problem is equivalent to the problem of minimizing

$$\sum_{y^n \in \{0,1\}^n} m(y^n, U),$$

as a function of U. Order now the elements of $\{P_i \beta^j \mid 1 \leq i \leq N;\ 0 \leq j \leq n\}$ in increasing order and denote them by $\alpha_1, \ldots, \alpha_{(n+1)N}$, $N = 2^n$.

We can write

$$\sum_{y^n \in \{1,0\}^n} m(y^n, U) = \sum_{l=1}^{(n+1)N} \alpha_l \, |\delta_l(U)|,$$

where $\delta_l(U) = \{y^n | m(y^n, U) = \alpha_l\}$. Further, set $\delta_l^*(U) = \delta_l(U) \cup \delta_{l+1}(U) \cup \ldots \ldots \cup \delta_{n \cdot N}(U)$. Then with $\alpha_0 := 0$

$$\sum_{l=1}^{(n+1)N} \alpha_l \, |\delta_l(U)| = \sum_{l=1}^{(n+1)N} (\alpha_l - \alpha_{l-1}) \, |\delta_l^*(U)|.$$

Since $\alpha_1 \geq 0$ and $\alpha_l - \alpha_{l-1} \geq 0$ for $l = 2, \ldots, (n+1)N$ we are done if the *same* U minimizes all $\delta_l^*(U)$; $l = 1, \ldots, (n+1)N$. We write now $\delta_l^*(U)$ as a union of spheres. Define radii

$$r_{li} = \begin{cases} -1 & \text{if } P_i < \alpha_l \\ \max\{t \mid t \text{ integer with } P_i \beta^t \geq \alpha_l\} & \text{else} \end{cases}$$

and observe that with the convention $S_{-1}(x^n) = \emptyset$

$$\delta_l^*(U) = \bigcup_{i=1}^{N} S_{r_{li}}(u_i).$$

Since $r_{l1} \geq r_{l2} \geq \ldots r_{lN}$ for $l = 1, \ldots, nN$ the general isoperimetry theorem gives the result.

References

1. *Shannon, C. E.*, Communication theory of secrecy systems, BSTJ, **28**, 656–715, 1949. The material in this paper appeared originally in a confidential report "A mathematical theory of cryptography" dated Sept. 1, 1945.
2. *Hellman, M. E.*, The Shannon theory approach to cryptography, IEEE Transactions on Inf. Theory, Oct. 1975.
3. *Harper, L. H.*, Optimal numberings and isoperimetric problems on graphs. Journal of Combinatorial Theory **1**, 385–393, 1966.
   Minimal numberings and isoperimetric problems on cubes. Theory of Graphs, International Symposium, Rome, ICC Dunod, 151–152, 1966.
4. *Hellman, M. E.*, An extension of the Shannon theory approach to cryptography. IEEE Transactions on Inf. Theory, **23**, no. *3*, 289–294, 1977.
5. *Lu, S. C.*, Bounds on Secrecy Systems, Presented at Intern. Symp. on Inf. Theory, Ithaca, N.Y. Oct. 10–14, 1977.

Замечания о шенноновских секретных системах связи

Р. АЛСВЕДЕ

(Биэлэфэлд)

Статья содердит три улучшения шенноновской теории секретных систем [1]:

1. С помощью очень простой конструкции определяются щифры, которые для естественных критериев безопасности столь же хороши, как и шенноновские "случайные шифры".

2. Для этой конструкции необязательно предполагать, что сообщения равновероятно. Шеннон сделал это предположение для того, чтобы использовать свой подход "случайного шифра".

3. Более того, конструируются оптимальные шифры при более грубых предположениях, что пользователям известно только ограничение на энтропию источника и что криптографу точно известна статистика сообщения.

Наконец, конструируются наихудшие коды для двоичного симметричного канала и подчеркивается значение этой "дуальной проблемы кодирования" для криптографии.

R. Ahlswede
Fakultät für Mathematik
der Universität Bielefeld
Universitätsstrasse 1
4800 Biel efeld 1