Problems of Control and Information Theory, Vol. 11, No. 5 (1982), pp. 337–351

BAD CODES ARE GOOD CIPHERS

R. AHLSWEDE, G. DUECK

(Bielefeld)

(Received August 10, 1981)

The paper investigates secrecy systems with additive-like instantaneous block encipherers subject to the error probability criterion. It is first shown that good ciphers for this system are bad codes for an associated discrete memoryless channel (DMC). Then asymptotically worst channel codes for any DMC are constructed explicitly. Applied to the secrecy system, this result gives an asymptotically optimal solution for the enciphering problem.

1. Introduction

In [1] C. E. Shannon presented a mathematical theory of secrecy systems. A good exposition and also certain extensions of this approach to cryptography were given by Hellman [2]. Ahlswede [4] proposed several improvements of Shannon's model and the problem of finding worst codes for the binary symmetrical channel.

Whereas these results are mainly concerned with abstract (block free) message sources, Lu [5], [6] considered in recent work secrecy systems with additive-like instantaneous block encipherers. Using random ciphering, Lu derived for these systems upper bounds on the probability of correct decryptment.

The main contribution of the present paper is the observation that to any additive-like instantaneous block encipherer one can associate a block code for a certain discrete memoryless channel, and that the best encipherer is obtained as the worst code for the associated channel. Hence it is now fully understood that the dual coding problem (the problem of finding worst codes), which in [4] was studied for its own sake and was felt to be relevant for the enciphering problem, is indeed the central problem of cryptography. In [4] the dual coding problem is solved exactly for the binary symmetric channel. For more general discrete memoryless channels this is still not the case, however, we give here an asymptotically optimal solution: asymptotically worst codes (and hence also asymptotically best sets of ciphers) are products of full sets of typical sequences and therefore asymptotically best ciphers are "clouds of sequences". Of course such sets of ciphers can be given explicitly.

It turns out that the connection between worst codes and best ciphers can be found also in more general systems as for instance those studied in [6]. As a result of this phenomenon we are able to give asymptotically *tight* bounds on the probability of correct decryptment for additive-like instantaneous block encipherers also in the case that side information is available to the "enemy".

Furthermore, it is a routine matter to derive a similar result for a more robust model (arbitrarily varying sources in the sense of [7]) and we therefore just state the result.

2. Worst codes are best ciphers

A. *The model*

We give first the model of ALIB encipherers, which was introduced by Lu ([5], [6]).

Let $\mathcal{C}$, $\mathcal{M}$, $\mathcal{S}$, $\mathcal{Y}$ be finite sets with

$$|\mathcal{C}| = |\mathcal{M}| = |\mathcal{Y}|.$$

We are given two correlated message sources $\{(X_i, S_i)\}_{i=1}^{\infty}$, where all the (X_i, S_i), $i = 1, 2, \ldots$ are independent replicae of a pair (X, S) of random variables with values in $\mathcal{M}\times\mathcal{S}$. We write $X^n = (X_1, \ldots, X_n)$ and $S^n = (S_1, \ldots, S_n)$. The joint distribution of (X^n, S^n) is given by

$$\Pr(X^n = m^n, S^n = s^n) = \prod_{i=1}^{n} \Pr(X = m_i, S = s_i)$$

for all $m^n = (m_1, \ldots, m_n) \in \mathcal{M}^n$, $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$. Furthermore, a combiner f is given, i.e. a function $f : \mathcal{C}\times\mathcal{M} \to \mathcal{Y}$, where $f(\cdot, m)$ is bijective for each $m \in \mathcal{M}$ and where $f(c, \cdot)$ is bijective for each $c \in \mathcal{C}$. $f^n : \mathcal{C}^n\times\mathcal{M}^n \to \mathcal{Y}^n$ denotes the n-fold product of f.

An (n, R) additive-like instantaneous block (ALIB) encipherer is a subset

$$\mathcal{E} \subset \mathcal{C}^n \text{ with } |\mathcal{E}| \leq \exp\{nR\}.$$

Give an (n, R) ALIB encipherer $\mathcal{E} \subset \mathcal{C}^n$ the system works as follows.

A *key word* $c^n \in \mathcal{E}$ is chosen according to the equidistribution on $\mathcal{E}$. The *encipherer* knows the *message word* m^n, an outcome of the random variable X^n, which is to be enciphered and to be sent to the *decipherer*. From the key word c^n selected and the message word m^n the encipherer forms the *cryptogram* $y^n = f^n(c^n, m^n)$ and sends it to the decipherer. The key word c^n is given

to the decipherer separately over a secure channel (a courier for instance). Knowing the key word c^n and the cryptogram y^n the decipherer deciphers the message word m^n, which is always possible because f is one-to-one.

The *cryptanalyst* intercepts the cryptogram y^n; he has also available the message word s^n, an outcome of the random variable S^n. S^n represents a side information source for the cryptanalyst. From the knowledge of y^n and s^n the cryptanalyst attempts to *decrypt* m^n (where we assume, that the cryptanalyst knows the "whole system", i.e. he knows the distribution of (X, S), $f, \mathcal{E}, n$). Since the cryptanalyst does not know the actual key word c^n being used he may decrypt a message $\bar{m}^n$ different from the true message word m^n being sent (error event). The error probability of the cryptanalyst is our measure for the quality of the chosen encipherer $\mathcal{E}$.

We now give a formal definition of $\lambda_c(\mathcal{E})$, the *probability* that the cryptanalyst *decrypts correctly*. Fix $\mathcal{E} \subset \mathcal{C}^n$. Let C^n be a random variable equidistributed on $\mathcal{E}$ and independent of X^n, S^n. Let $Y^n = f^n(C^n, X^n)$. If the cryptanalyst intercepts y^n and gets the side information s^n, he can maximize the probability of decrypting correctly by deciding on a message word $\bar{m}^n$ for which

$$\Pr(X^n = \bar{m}^n \mid S^n = s^n,\ Y^n = y^n) = \max_{m^n} \{\Pr(X^n = m^n \mid S^n = s^n,\ Y^n = y^n)\}.$$

Since this maximum likelihood decision rule is optimal in the error probability sense, we always assume that the cryptanalyst uses this rule. Therefore, we define for an ALIB encipherer $\mathcal{E} \subset \mathcal{C}^n$:

$$\lambda_c(\mathcal{E}) = \sum_{s^n \in \mathcal{S}^n} \sum_{y^n \in \mathcal{Y}^n} \Pr(S^n = s^n,\ Y^n = y^n) \max_{m^n} \{\Pr(X^n = m^n \mid S^n = s^n,\ Y^n = y^n)\} =$$

$$= \sum_{s^n \in \mathcal{S}^n} \sum_{y^n \in \mathcal{Y}^n} \max_{m^n} \{\Pr(X^n = m^n,\ S^n = s^n,\ Y^n = y^n)\}. \tag{2.1}$$

We note here that the definition of the combiner f implies that each of the random variables C^n, X^n, Y^n is a function of the remaining two others. Therefore the cryptanalyst also could try to decrypt the true key word c^n correctly and then he could compute m^n from c^n and y^n.

This leads to the observation that $\lambda_c(\mathcal{E})$ can also be expressed as

$$\lambda_c(\mathcal{E}) = \sum_{s^n \in \mathcal{S}^n} \sum_{y^n \in \mathcal{Y}^n} \max_{c^n \in \mathcal{E}} \{\Pr(C^n = c^n,\ S^n = s^n,\ Y^n = y^n)\}. \tag{2.2}$$

We define for any n and $R > 0$:

$$\lambda_c(n, R) = \min_{\mathcal{E}} \lambda_c(\mathcal{E}),$$

where the minimum is taken over all ALIB encipherers $\mathcal{E} \subset \mathcal{C}^n$ with $|\mathcal{E}| \leq \leq \exp\{nR\}$. Our aim is to derive a computable expression for

$$-\lim \frac{1}{n} \log \lambda_c(n, R).$$

B. *Statement of the results and connection to coding theory*

For random variables $\bar{Z}, Z$ with values in a finite set $\mathcal{Z}$ $H(Z)$ denotes the familiar entropy function, and

$$D(\bar{Z} \| Z) = \sum_{z \in \mathcal{Z}} \Pr(\bar{Z} = z) \log \frac{\Pr(\bar{Z} = z)}{\Pr(Z = z)}$$

stands for the informational I-divergence.

For triples $(C, \tilde{X}, \tilde{S})$ of random variables with values in $\mathcal{C} \times \mathcal{M} \times \mathcal{S}$ set

$$F(C, \tilde{X}, \tilde{S}) = D((\tilde{X}, \tilde{S}) \| (X, S)) + H(C | f(C, \tilde{X}), \tilde{S}).$$

(We shall write in the sequel $D(\tilde{X}, \tilde{S} \| X, S)$ instead of $D((\tilde{X}, \tilde{S}) \| (\tilde{X}, \tilde{S}))$.) For $R > 0$ we define

$$F^*(R) = \max_{\substack{C : H(C) \leq R \\ C \text{ independent of } (X,S)}} \min_{\tilde{X}, \tilde{S}} F(C, \tilde{X}, \tilde{S}).$$

Let F^{**} be the concave upper envelope of F^*. We shall prove:

Theorem 1. For any $R > 0$

$$-\lim_{n \to \infty} \frac{1}{n} \log \lambda_c(n, R) = F^{**}(R).$$

We establish now a connection between the present problem and the problem to find the worst code for a discrete memoryless channel.

A discrete memoryless channel consists of a finite input alphabet $\mathcal{U}$, a finite output alphabet $\mathcal{V}$, and a set of transmission probabilities $\{W(v|u) | u \in \mathcal{U}, v \in \mathcal{V}\}$.

The transmission probabilities for n-words

$$u^n = (u_1, \ldots, u_n) \in \mathcal{U}^n, \; v^n = (v_1, \ldots, v_n) \in \mathcal{V}^n$$

are given by

$$W^n(v^n | u^n) = \prod_{i=1}^{n} W(v_i | u_i).$$

An (n, R) code for W is a subset $\mathcal{E} \subset \mathcal{U}^n$ with $|\mathcal{E}| \leq \exp\{nR\}$. The probability $\bar{\lambda}_c(\mathcal{E})$ of correct decoding (maximum-likelihood) is

$$\bar{\lambda}_c(\mathcal{E}) = \sum_{v^n} \Pr(V^n = v^n) \max_{u^n \in \mathcal{E}} \Pr(U^n = u^n \mid V^n = v^n).$$

Here, U^n is a random variable equidistributed on $\mathcal{E}$ and V^n is the corresponding output variable with respect to W (we say also that "W connects U^n and V^n"), i.e.

$$\Pr(V^n = v^n) = \sum_{u^n} \Pr(U^n = u^n) W^n(v^n \mid u^n).$$

After these definitions we return to the ALIB enciphering problem. Let $\mathcal{E} \subset \mathcal{C}^n$ be an (n, R) ALIB encipherer. We show that $\mathcal{E}$ can be viewed as a code for a special discrete memoryless channel.

Let C^n be equidistributed on $\mathcal{E}$ and independent of (X^n, S^n); $Y^n = f^n(C^n, X^n)$. Then for any $c^n \in \mathcal{E}$:

$$\Pr(Y^n = y^n, S^n = s^n \mid C^n = c^n) =$$

$$= \sum_{m^n} \Pr(X^n = m^n) \Pr(Y^n = y^n, S^n = s^n \mid C^n = c^n, X^n = m^n),$$

because C^n and X^n are independent. Observe that by the definition of the source

$$\Pr(X^n = m^n) = \prod_{i=1}^{n} \Pr(X = m_i)$$

and

$$\Pr(Y^n = y^n, S^n = s^n \mid C^n = c^n, X^n = m^n) =$$

$$= \Pr(S^n = s^n \mid X^n = m^n) \cdot \Pr(Y^n = y^n \mid C^n = c^n, X^n = m^n),$$

because C^n is independent of (X^n, S^n) and $Y^n = f^n(C^n, X^n)$.

Obviously

$$\Pr(S^n = s^n \mid X^n = m^n) = \prod_{i=1}^{n} \Pr(S = s_i \mid X = m_i)$$

and

$$\Pr(Y^n = y^n \mid C^n = c^n, X^n = m^n) = \prod_{i=1}^{n} \delta(y_i, f(c_i, m_i)),$$

where $\delta(y_i, f(c_i, m_i))$ equals 1 if $y_i = f(c_i, m_i)$ and $\delta(y_i, f(c_i, m_i)) = 0$ otherwise.

Thus

$$\Pr(Y^n = y^n, S^n = s^n \mid C^n = c^n) = \tag{2.3}$$

$$= \prod_{i=1}^{n} \sum_{m_i \in \mathfrak{M}} \Pr(X = m_i)\Pr(S = s_i | X = m_i) \cdot \delta(y_i, f(c_i, m_i)) .$$

If we define now a channel W^* with input alphabet $\mathfrak{C}$ and output alphabet $\mathfrak{Y} \times \mathfrak{S}$ by

$$W^*(y, s|c) = \sum_{m \in \mathfrak{M}} \Pr(X = m) \cdot \Pr(S = s | X = m) \cdot \delta(y, f(c, m)),$$

then we see that equation (2.3) can be written as

$$\Pr(Y^n = y^n, S^n = s^n | C^n = c^n) = \prod_{i=1}^{n} W^*(x_i, s_i | c_i) . \tag{2.4}$$

From (2.2), (2.4), and the deinition of a channel code and its probability of correct decoding *we observe now that the* (n, R) *ALIB encipherer* $\mathfrak{E}$ *can be viewed as an* (n, R) *block code for the discrete memoryless channel* W^*. It is obvious that the probability $\lambda_c(\mathfrak{E})$ of correct decryptment for the encipherer $\mathfrak{E}$ is equal to the probability $\bar{\lambda}_c(\mathfrak{E})$ of correct decoding for the *code* $\mathfrak{E}$ and channel W^*.

Hence the problem to find the best ALIB encipherer is a special case of the dual coding problem, i.e., the problem to find worst codes for discrete memoryless channels. We give here an asymptotic solution. For the special case of the binary symmetric channel (BSC), Ahlswede [4] gave an exact solution. Worst codes for the BSC are quasi-Hamming spheres in $\{0, 1\}^n$. This result can be derived via the isoperimetry theorem of Harper [3]. In order to formulate the asymptotic solution of the dual coding problem for general discrete memoryless channels we need a few definitions.

We shall define the functions E, E^*, E^{**} as the analogues to F, F^*, F^{**}:

Consider the channel given by $\mathfrak{U}, \mathfrak{V}$, and W. For a random variable U with values in $\mathfrak{U}$ and another (dummy) channel $\tilde{W}$ between $\mathfrak{U}$ and $\mathfrak{V}$ we define

$$D(\tilde{W} || W | U) = \sum_{u} \Pr(U = u) \sum_{v} \tilde{W}(v|u) \log \frac{\tilde{W}(v|u)}{W(v|u)} .$$

Further $\tilde{V}$ denotes the output variable corresponding to U with respect to $\tilde{W}$.

We define

$$E(U, \tilde{W}) = D(\tilde{W} || W | U) + H(U | \tilde{V})$$

and for $R > 0$

$$E^*(R) := \max_{U: H(U) \leq R} \min_{\tilde{W}} E(U, \tilde{W}).$$

Finally let E^{**} be the concave upper envelope of E^*. Let $\bar{\lambda}_c(n, R) = \min \bar{\lambda}_c(\mathcal{C})$, where the minimum is taken over all (n, R) codes $\mathcal{C}$.

Theorem 2. For any $R > 0$

$$-\lim_{n\to\infty} \frac{1}{n} \log \bar{\lambda}_c(n, R) = E^{**}(R).$$

Since Theorem 2 implies Theorem 1 we have to prove only Theorem 2. It is a routine matter to verify that the functions F, F^*, F^{**} are specialized versions of E, E^*, E^{**}.

The proof of Theorem 2 has two parts: the converse part and the direct part. From the proof of the direct part it can be seen that the worst codes (and therefore the best encipherers) are products of full sets of typical sequences. This is in coincidence with the result of [4], where it is shown that worst codes for the BSC consist of quasi-spheres in the Hamming space $\{0, 1\}^n$.

3. Proof of Theorem 2

We introduce some notation. *The type* of a sequence $u^n \in \mathcal{U}^n$ is the distribution P_{u^n} on $\mathcal{U}$ defined by letting $P_{u^n}(u)$ count the relative frequency of the letter u in the n-sequence u^n. *Joint types* such as the type $P_{(u^n,v^n)}$ of $(u^n, v^n) \in \in \mathcal{U}^n \times \mathcal{V}^n$ are defined analogously. For a distribution P on $\mathcal{U}$ the set of all P-typical sequences in $\mathcal{U}^n$ is defined by

$$\mathcal{T}_P^n := \left\{u^n \middle| |P_{u^n}(u) - P(u)| \leq \frac{P(u) \log n}{\sqrt{n}} \text{ for all } u \in \mathcal{U}\right\}.$$

We state some well-known properties of typical sequences:

Lemma 1

a) Let P be a distribution on $\mathcal{U}$ and P^n its n-th extension. Then

$$P^n(\mathcal{T}_P^n) \geq 1 - \frac{\varkappa}{\log n}$$

for some constant $\varkappa$ depending only on $|\mathcal{U}|$.

b) $P^n(u^n) = \exp\{-n(D(P_{u^n}||P) + H(P_{u^n}))\}$
for all $u^n \in \mathcal{U}^n$.

c) Let U be a random variable on $\mathcal{U}$ and $\mathcal{W}$ a channel between $\mathcal{U}$ and $\mathcal{V}$. V denotes the output variable to U with respect to $\mathcal{W}$. Let P be the joint

distribution of the pair $(U, \bar{V})$. Then for any pair

$$(u^n, v^n) \in \mathcal{S}_P^n \subset \mathcal{U}^n \times \mathcal{V}^n$$

$$W^n(v^n|u^n) = \exp\{-n(D(\bar{W}||W|U) + H(\bar{V}|U) + o(n))\}.$$

A. *Proof of the converse part of Theorem 2*

Let $\mathcal{E} \subset \mathcal{U}^n$ be an (n, R) code.

Since it is obviously sufficient to prove Theorem 2 for code words of fixed composition we assume that there is a distribution Q on $\mathcal{U}$ such that

$$P_{u^n} = Q \quad \text{for any} \quad u^n \in \mathcal{E}.$$

Let U denote a random variable on $\mathcal{U}$ with distribution Q. Let $U^n = (U_1, \ldots, U_n)$ be a random variable equidistributed on $\mathcal{E}$. Let V^n denote the output variable corresponding to U^n with respect to W. For convenience we make the convention that $\mathcal{U}^0$ is a set of cardinality one; U_0 is a constant variable with value in $\mathcal{U}^0$. For the given code $\mathcal{E}$ we shall now estimate $\bar{\lambda}_c(\mathcal{E})$. We shall proceed in this way:

$$(3.1)\quad \bar{\lambda}_c(\mathcal{E}) = \sum_{\tilde{Q} \text{ distr. on } \mathcal{U}\times\mathcal{V}} \Pr((U^n, V^n) \text{ has type } Q) \cdot \Pr(\text{correct decoding} \mid (U^n, V^n) \text{ has type } \tilde{Q}) \geq \Pr((U^n, V^n) \text{ has type } \tilde{Q}) \cdot \Pr(\text{correct decoding} \mid (U^n, V^n) \text{ has type } \tilde{Q}),$$

where $\tilde{Q}$ is any distribution on $\mathcal{U}\times\mathcal{V}$.

We define first a distribution P on $\mathcal{U}\times\mathcal{V}$, for which the estimate in (3.1) becomes sufficiently sharp. Then we relate the probabilities in (3.1) to I-divergences and information quantities. Finally we shall give a single-letterization of these quantities. For any $i = 1, \ldots, n$ and any $u^{i-1} \in \mathcal{U}^{i-1}$ let $U(u^{i-1})$ be a random variable with values in $\mathcal{U}$ such that

$$(3.2)\qquad \Pr(U(u^{i-1}) = u) := \Pr(U_i = u \mid U^{i-1} = u^{i-1}).$$

For any $i = 1, \ldots, n$ and any $u^{i-1} \in \mathcal{U}^{i-1}$ let $V(u^{i-1})$ be a random variable with values in $\mathcal{V}$ and let $W(u^{i-1})$ be a "dummy" channel between $\mathcal{U}$ and $\mathcal{V}$ such that $V(u^{i-1})$ is the corresponding output variable to $U(u^{i-1})$ with respect to the channel $W(u^{i-1})$ and such that

$$(3.3)\qquad E(U(u^{i-1}), W(u^{i-1})) = \min_{\bar{W}} E(U(u^{i-1}), \bar{W}),$$

where the min is taken over all channels $\bar{W}$ between $\mathcal{U}$ and $\mathcal{V}$.

Now we define the distribution P on $\mathcal{U}\times\mathcal{V}$ by

(3.4)

$$P(u, v) := \frac{1}{n}\sum_{i=1}^{n}\sum_{u^{i-1}\in\mathcal{U}^{i-1}} \Pr(U^{i-1} = u^{i-1}) \cdot \Pr(U(u^{i-1}) = u,\ V(u^{i-1}) = v).$$

Since by (3.2) and (3.4) for any $u \in \mathcal{U}$

$$\sum_{v} P(u, v) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u^{i-1}\in\mathcal{U}^{i-1}} \Pr(U^{i-1} = u^{i-1}) \cdot \Pr(U_i = u | U^{i-1} = u^{i-1})$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Pr(U_i = u) = \Pr(U = u) = Q(u),$$

the marginal distribution of P on $\mathcal{U}$ is the distribution Q of the given random variable U which is generated by $\mathcal{E}$. Therefore it makes sense to define a random variable $\tilde{V}$ on $\mathcal{V}$ and a channel $\tilde{W}$ between $\mathcal{U}$ and $\mathcal{V}$ such that $(U, \tilde{V})$ has distribution P and such that U and $\tilde{V}$ are connected by $\tilde{W}$.

Now we are going to estimate $\bar{\lambda}_c(\mathcal{E})$ for the given code $\mathcal{E}$. In order to be able to apply Lemma 1 we rather look at the k-fold product $\mathcal{E}^k \subset \mathcal{U}^{nk}$, where k is a positive integer. Since $\mathcal{E}$ is an (n, R) code, $\mathcal{E}^k$ is an (nk, R) code. Let U^{nk} (resp. V^{nk}) be the k-fold product of U^n (resp. V^n). Note that U^{nk} is equidistributed on $\mathcal{E}^k$.

The probability $\bar{\lambda}_c(\mathcal{E}^k)$ is given by

(3.5) $$\bar{\lambda}_c(\mathcal{E}^k) = \sum_{v^{nk}\in\mathcal{V}^{nk}} \max_{u^{nk}\in\mathcal{U}^{nk}} \{\Pr(U^{nk} = u^{nk},\ V^{nk} = v^{nk})\}.$$

By the product structure of $\mathcal{E}^k$ we have for

$$v^{nk} = (v_1^n, \ldots, v_k^n) \in \mathcal{V}^{nk}$$

$$\max_{u^{nk}=(u_1^n,\ldots,u_k^n)} \{\Pr(U^{nk} = u^{nk},\ V^{nk} = v^{nk})\} = \prod_{i=1}^{k} \max_{u_i^n\in\mathcal{U}^n} \{\Pr(U^n = u_i^n,\ V^n = v_i^n)\}.$$

Hence we can conclude

$$\bar{\lambda}_c(\mathcal{E}^k) = (\bar{\lambda}_c(\mathcal{E}))^k.$$

We shall now estimate $\bar{\lambda}_c(\mathcal{E}^k)$ rather than $\bar{\lambda}_c(\mathcal{E})$.

Let $\mathcal{A}^k = \{v^{nk} \in \mathcal{V}^{nk} |$ there is a $u^{nk} \in \mathcal{E}^k$ such that $(u^{nk}, v^{nk}) \in \mathcal{T}_P^{nk}\}$, where $\mathcal{T}_P^{nk} \subset \mathcal{U}^{nk}\times\mathcal{V}^{nk}$ is the set of all jointly P-typical sequences.

Thus, since $\mathcal{U}^{nk}$ is equidistributed on $\mathcal{E}^k$,

(3.6) $$\bar{\lambda}_c(\mathcal{E}^k) \geq |\mathcal{E}|^{-k} \sum_{v^{nk}\in\mathcal{A}^k} \max_{u^{nk}\in\mathcal{E}^k:(u^{nk},v^{nk})\in\mathcal{T}_P^{nk}} \{\Pr(V^{nk} = v^{nk} | U^{nk} = u^{nk})\}.$$

We apply Lemma 1 to the distribution P of $(U, \tilde{V})$ and get by the definitions of $(U, \tilde{V})$ and $\tilde{W}$

$$\Pr(V^{nk} = v^{nk} | U^{nk} = u^{nk}) \geq \exp\{-nk\, D(\tilde{W}||W|U) - nkH(\tilde{V}|U) - n \cdot o(k)\}$$

for any pair $(u^{nk}, v^{nk}) \in \mathcal{T}_P^{nk}$.

Hence, we can derive from (3.6)

$$(3.7)\quad \bar{\lambda}_c(\mathcal{C}^k) \geq |\mathcal{C}|^{-k} \exp\{-nk(D(\tilde{W}||W|U) + H(\tilde{V}|U)) - n \cdot o(k)\} \cdot |\mathcal{A}^k|.$$

As a second step in the proof we have to relate $|\mathcal{A}|^k$ to information quantities.

For this purpose we define the random variable $\tilde{V}^n$ on $\mathcal{V}^n$ by

$$(3.8)\quad \Pr(\tilde{V}^n = v^n | U^n = u^n) = \prod_{i=1}^{n} \Pr(V(u^{i-1}) = v_i | U(u^{i-1}) = u_i)$$

for all $u^n = (u_1, \ldots, u_n) \in \mathcal{U}^n$, $v^n = (v_1, \ldots, v_n) \in \mathcal{V}^n$, where $V(u^{i-1})$, $U(u^{i-1})$ are the random variables defined in (3.2) and (3.3).

By the definition of the channels $W(u^{i-1})$,

$$(3.9)\quad \Pr(\tilde{V}^n = v^n | U^n = u^n) = \prod_{i=1}^{n} W(u^{i-1})(V(u^{i-1}) = v_i | U(u^{i-1}) = u_i).$$

We write $\tilde{V}^n = (\tilde{V}_1, \ldots, \tilde{V}_n)$. $\tilde{V}^{nk}$ is the k-product of $\tilde{V}^n$. From (3.9) and the definition of P in (3.4) we observe that for any $(u, v) \in \mathcal{U} \times \mathcal{V}$

$$\frac{1}{n} \sum_{i=1}^{n} \Pr(\tilde{V}_i = v, U_i = u) = P(u, v).$$

Therefore one can easily see that by Lemma 1

$$\Pr((\tilde{V}^{nk}, U^{nk}) \in \mathcal{T}_P^{nk}) \geq 1 - \frac{\varkappa}{\log k}$$

and thus

$$(3.10)\quad \Pr(\tilde{V}^{nk} \in \mathcal{A}^k) \geq 1 - \frac{\varkappa}{\log k}.$$

Inequality (3.10) enables us to compare $\log |\mathcal{A}^k|$ with $H(\tilde{V}^{nk})$. In fact, the grouping axiom for the entropy function and (3.10) yield

$$(3.11)\quad H(\tilde{V}^{nk}) \leq \log 2 + \log|\mathcal{A}^k| + \frac{\varkappa}{\log k} \log |\mathcal{V}^{nk}|.$$

Taking (3.11) into account we get from (3.7)

$$\bar{\lambda}_c(\mathcal{E}^k) \geq |\mathcal{E}|^{-k} \cdot \exp\{-nk(D(\tilde{W}||W|U) + H((\tilde{V}|U)) - \\ - no(k)\} \cdot \exp\{H(\tilde{V}^{nk})\}. \tag{3.12}$$

For the remainder of the proof of the converse part we have to give a single letter expression for the right-hand side of (3.12).

First note that

$$H(\tilde{V}^{nk}) = kH(\tilde{V}^n),$$

because of the product structure of $\tilde{V}^{nk}$. Further,

$$H(\tilde{V}^n) = \sum_{i=1}^{n} H(\tilde{V}_i|\tilde{V}^{i-1}) \geq \sum_{i=1}^{n} H(\tilde{V}_i|\tilde{V}^{i-1}U^{i-1}) \\ = \sum_{i=1}^{n} H(\tilde{V}_i|U^{i-1}),$$

because $\tilde{V}_i \to U^{i-1} \to \tilde{V}^{i-1}$ form a Markov chain in this order. This follows from (3.8).

From (3.8) we observe that

$$\sum_{i=1}^{n} H(\tilde{V}_i|U^{i-1}) = \sum_{i=1}^{n} \sum_{u^{i-1} \in \mathcal{U}^{i-1}} \Pr(U^{i-1} = u^{i-1})\, H(V(u^{i-1})). \tag{3.13}$$

Of course we also have

$$\log |\mathcal{E}| = H(U^n) = \sum_{i=1}^{n} H(U_i|U^{i-1}) = \\ = \sum_{i=1}^{n} \sum_{u^{i-1} \in \mathcal{U}^{i-1}} \Pr(U^{i-1} = u^{i-1})\, H(U(u^{i-1})). \tag{3.14}$$

From (3.12) we see that we have now to split up $D(\tilde{W}||W|U) + H(\tilde{V}|U)$ in an appropriate way. This is easy (use (3.4)):

$$D(\tilde{W}||W|U) + H(\tilde{V}|U) = -\sum_{u} \Pr(U = u) \sum_{v} \tilde{W}(v|u) \log W(v|u) \\ = -\sum_{u,v} P(u, v) \log W(v|u) \\ = -\frac{1}{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{u^{i-1} \in \mathcal{U}^{i-1}} \Pr(U^{i-1} = u^{i-1}) \cdot \\ \cdot \sum_{u,v} \Pr(U(u^{i-1}) = u, V(u^{i-1}) = v) \cdot \log W(v|u) \\ = \frac{1}{n} \sum_{i=1}^{n} \sum_{u^{i-1} \in \mathcal{U}^{i-1}} \Pr(U^{i-1} = u^{i-1}) \cdot \\ \cdot [D(W(u^{i-1})||W|U(u^{i-1})) + H(V(u^{i-1})|U(u^{i-1}))]. \tag{3.15}$$

This, (3.12), (3.13), and (3.14) imply

$$\bar{\lambda}_c(\mathcal{E}^k) \geq \exp\{-k\,(\sum_{i=1}^{n} \sum_{u^{i-1}\in\mathcal{U}^{i-1}} \Pr(U^{i-1}=u^{i-1})\,E(U(u^{i-1}),\,W(u^{i-1}))) - no(k)\}.$$

Now use (3.3) and the definition of E^* and E^{**} to obtain $\bar{\lambda}_c(\mathcal{E}^k) \geq$

$$\geq \exp\left\{-k\,E^{**}\left(\frac{1}{n}\log|\mathcal{E}|\right) - no(k)\right\}, \text{ which completes the proof.}$$

B. *Proof of the direct part of Theorem 2*

Fix a block length n and a distribution Q on $\mathcal{U}$ such that

$$Q(u) \in \{0,\, 1/n,\, 2/n,\, \ldots,\, 1\} \quad \text{for } u \in \mathcal{U}.$$

We define the code

$$\mathcal{E} = \{u^n \in \mathcal{U}^n | P_{u^n} = Q\}. \tag{3.16}$$

Let U be a random variable on $\mathcal{U}$ with distribution Q. We shall prove:

$$\bar{\lambda}_c(\mathcal{E}) \leq \exp\{-n \min_{\tilde{W}} E(U, \tilde{W}) + o(n)\}. \tag{3.17}$$

Note that if this would be proved, then the direct part of Theorem 2 follows by the standard time sharing argument. Therefore, asymptotically "worst" codes are products of full sets of typical sequences (of the form (3.16))

We give here only a bit more than an outline of the complete proof. Since $\mathcal{E}$ is defined to be the set of all sequences of type Q, $\mathcal{E}$ has "almost product structure". Therefore, the evaluation of $\bar{\lambda}_c(\mathcal{E})$ is very straightforward and needs only simple counting arguments for sets of typical sequences.

We start: Let U^n be equidistributed on $\mathcal{E}$. Let V^n be the output variable corresponding to U^n with respect to W. Let $\mathcal{S}_n = \mathcal{S}_n(\mathcal{U}, \mathcal{V})$ be the set of all distributions P on $\mathcal{U}\times\mathcal{V}$ satisfying

$$P(u, v) \in \{0,\, 1/n,\, 2/n,\, \ldots,\, 1\}$$

for any $(u, v) \in \mathcal{U}\times\mathcal{V}$. It is well known that

$$|\mathcal{S}_n| \leq (n+1)^{|\mathcal{U}|\cdot|\mathcal{V}|}. \tag{3.18}$$

We estimate $\bar{\lambda}_c(\mathcal{E})$ as follows:

$$\bar{\lambda}_c(\mathcal{E}) = \sum_{v^n} \max_{u^n} \{\Pr(U^n = u^n,\, V^n = v^n)\}$$

$$\leq \sum_{v^n} \sum_{P\in\mathcal{S}_n} \max_{\substack{u^n:(u^n,v^n)\\ \text{has type } P}} \{\Pr(U^n = u^n,\, V^n = v^n)\}.$$

Since $\mathcal{E} = \{u^n | P_{u^n} = Q\}$ and since U has distribution Q,

$$|\mathcal{E}|^{-1} \leq \exp\{-nH(U) + o(n)\}. \tag{3.19}$$

Since U^n is equidistributed on $\mathcal{E}$, we can conclude for (u^n, v^n) with $P_{(u^n, v^n)} = P$, $u^n \in \mathcal{E}$ that

$$\Pr(V^n = v^n | U^n = u^n) \leq \exp\{-n(D(\tilde{W}_p || W|U) + H(\tilde{V}_p|U)) + o(n)\}, \tag{3.20}$$

where $\tilde{V}_p$ is a random variable on $\mathcal{Y}$ such that $(U, \tilde{V}_p)$ has distribution P and where $\tilde{W}_p$ is the channel connecting U and $\tilde{V}_p$ (Lemma 1).

In summary,

$$\bar{\lambda}_c(\mathcal{E}) \leq |\mathcal{S}_n| \cdot \max_{P \in \mathcal{S}_n} |\mathcal{B}(P)| \cdot \exp\{-n(D(\tilde{W}_p || W|U) + H(\tilde{V}_p|U) + H(U)) + o(n)\},$$

where

$$\mathcal{B}(P) := \{v^n \in \mathcal{Y}^n | \text{ there is a } u^n \text{ with } P_{(u^n, v^n)} = P\}.$$

It is easy to see that

$$|\mathcal{B}(P)| \leq \exp\{nH(\tilde{V}_p) + o(n)\}.$$

Since

$$H(\tilde{V}_p|U) + H(U) - H(\tilde{V}_p) = H(U|\tilde{V}_p),$$

we therefore have

$$\begin{aligned} \bar{\lambda}_c(\mathcal{E}) &\leq |\mathcal{S}_n| \cdot \max_{P \in \mathcal{S}_n} \exp\{-n(D(\tilde{W}_p || W|U) + H(U|\tilde{V}_p)) + o(n)\} \\ &\leq |\mathcal{S}_n| \cdot \exp\{-n \cdot \min_{\tilde{W}} E(U, \tilde{W}) + o(n)\}. \end{aligned}$$

This gives (3.17) by applying (3.18).

4. On a more robust model

We consider the former model (Section 2. A) with one modification for the sake of robustness. In Section 2 we were given two message sources $\{(X_l, S_l)\}_{l=1}^{\infty}$, where the X_l's represented the message sequence to be enciphered and where the S_l's served as side information for the "enemy". Now we are given a set $\mathcal{Q} = \{P(\cdot|s) | s \in \mathcal{S}\}$ of probability distributions on $\mathcal{M}$. For every $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$ define the distribution $P^n(\cdot|s^n)$ on $\mathcal{M}^n$ by

$$P^n(m^n|s^n) = \prod_{l=1}^{n} P(m_l|s_l), \quad m^n = (m_1, \ldots, m_n) \in \mathcal{M}^n.$$

The difference to a correlated source $\{(X_t, S_t)\}_{t=1}^{\infty}$ is that not the joint distribution but only the conditional distributions are specified and that the $\mathcal{S}$-output is not governed by a probabilistic law. A sequence $s_1, s_2, \ldots$ can be viewed as a "state sequence" of the source.

The definition of an ALIB encipherer $\mathcal{E} \subset \mathcal{C}^n$ is the same as in Section 2. However, since we have no probabilities on the s^n's, we have to change the criterion of error.

We look now for ciphers for which

$$\lambda_c^*(\mathcal{E}) = \max_{s^n} \sum_{y^n \in \mathcal{Y}^n} \max_{m^n} \{\Pr(X^n = m^n, Y^n = y^n | S^n = s^n)\}$$

becomes small (compare with (2.1)).

Again it is possible to show that the asymptotically best ALIB encipherers are products of full sets of typical sequences. This can be seen from the exponents given in Theorem 3. We shall omit the proof of this Theorem because it is only a translation of the proof of Theorem 1 into the new terms. We formulate the result:

For any random variable S on $\mathcal{S}$ let $X(S)$ be a random variable on $\mathcal{M}$ such that $\Pr(X(S) = m) = \sum_s \Pr(S = s) \cdot P(m|s)$. We denoted by

$$D(\tilde{X}||X(S)|S) = \sum_s \Pr(S = s) \sum_m \Pr(\tilde{X} = m|S = s) \log \frac{\Pr(\tilde{X}=m|S=s)}{\Pr(X(S)=m|S=s)}$$

the conditional I-divergence.

For triples $(C, \tilde{X}, S)$ of random variables on $\mathcal{C} \times \mathcal{M} \times \mathcal{S}$ set

$$\overline{F}(C, \tilde{X}, S) = D(\tilde{X}||X(S)|S) + H(C|f(C, \tilde{X}), S)$$

and for $R > 0$

$$\overline{F}^*(R) = \max_{\substack{C: H(C) \leq R \\ C \text{ independent of all } S.}} \min_{S, \tilde{X}} \overline{F}(C, \tilde{X}, S)$$

Theorem 3. For any $R > 0$

$$-\lim_{n \to \infty} \frac{1}{n} \log \lambda_c^*(n, R) = \overline{F}^{**}(R).$$

References

1. *Shannon, C. E.*, Communication theory of secrecy systems. BSTJ, **28**, 656–715, 1949.
2. *Hellman, M. E.*, An extension of the Shannon theory approach to cryptography, IEEE Trans. on Inf. Theory, **23**, 3, 289–294, 1977.
3. *Harper, L. H.*, Optimal numberings and isoperimetric problems on graphs, J. Combinatorial Theory, 385–393, 1966.

4. *Ahlswede, R.*, Remarks on secrecy systems, Presented at the International Symposium on Inf. Theory, Ithaca, N. Y., Oct. 10–14, 1977, to appear in Problems of Control and Information Theory.
5. *Lu, S. C.*, Random ciphering bounds on a class of secrecy systems and discrete message sources, IEEE Trans. Inf. Theory, **IT–25**, *4*, 405–414, 1979.
6. *Lu* S. C., Secrecy systems with side information about the message available to a cryptanalyst, IEEE Transactions on Inf. Theory, **25**, *4*, 472–475, 1979.
7. *Ahlswede, R.*, Coloring hypergraphs: a new approach to multi-user source coding, Part I, J. Combinatorics, Information & System Sciences, **4**, *1*, 76–115, 1979, Part II, ib. **5**, *3*, 202–268, 1980.

Плохие коды есть хорошие шифры

Р. АЛСВЕДЕ, Г. ДЮК

(Билефелд)

В работе рассматриваются системы засекречивания с аддитивными моментальными блочными устройствами засекречивания с точки зрения критерия по вероятности ошибок. Прежде всего показано, что хорошие шифры для такой системы являются плохими кодами для соответствующего дискретного канала без памяти. Сконструированы асимптотически худшие коды для любого дискретного канала без памяти. Применяя этот результат к системе засекречивания, получается асимптотически оптимальное решение проблемы шифрования.

R. Ahlswede, G. Dueck
Faculty of Mathematics
University of Bielefeld
D-4800 Bielefeld