Extension of Shannon's Theory of Ciphers based on Latin Rectangles

Karel Burda

Brno University of Technology, Brno, Czech Republic

Summary
The paper extends Shannon's classical theory of ciphers. Here ciphers are modeled by Latin rectangles and their resistance to brute force attack is assessed through the valence of cryptograms. The valence of a cryptogram is defined as the number of all meaningful messages produced by decrypting the cryptogram with all possible keys. In this paper, the mean cryptogram valence of an arbitrary modern cipher with K keys, N outputs and N inputs, of which M inputs are messages, is derived. Furthermore, the lower bound on the valence of the cryptograms of entire ciphers is derived in this paper. The obtained parameters allow to assess the resistance of cryptograms, resp. ciphers against brute force attack. The model is general, illustrative and uses a simpler mathematical apparatus than existing theory. Therefore, it can also be used as an introduction to the theory of resistance of ciphers to brute force attack.

Keywords:
Shannon, secrecy systems, brute force attack, Latin rectangles.

1. Introduction

People communicate with each other using messages, which are sequences of symbols in which are encoded information. The symbols used can be expressed in terms of numbers, and so each message can be represented as a unique number m. Encryption cryptosystems or ciphers are used to hide the contents of messages (see Fig. 1). On the sender's side, an encryption function E is used to assign a seemingly random number c to the input number m. This pseudo-random number is called a cryptogram. The function E is randomly selected from K of possible encryption functions using a parameter called the encryption key e. We will therefore formally write the encryption as $c = \mathrm{E}(m, e)$.

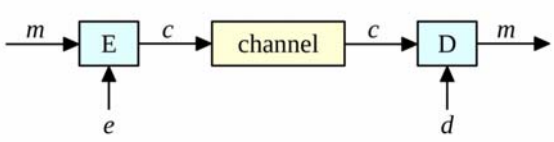


Fig. 1: Encryption cryptosystem

The sender sends the cryptogram c over the transmission channel to the counter-party. The recipient uses a secret parameter, called the decryption key d, to select from all K possible decryption functions this function D that is the inverse of the encryption function E used. Thus, the output of the decryption function will be the original message m. Formally, we express this in the notation $m = \mathrm{D}(c, d)$.

A possible attacker can intercept on the cryptogram c in the transmission channel. However, since he does not know the secret decryption key d, he cannot invert the cryptogram into the form of the message m. However, he can attempt to break the cryptogram, i.e., to discover the transmitted message without knowing the key. A universal method of breaking a cipher is the so-called brute force attack, during which the attacker decrypts the cryptogram with all possible keys while analyzing the meaningfulness of the obtained results. In this way, he may discover that the cryptogram c could have been created by encrypting any message from a total of v different messages. The larger the value of v, the greater the attacker's uncertainty about which of the v possible messages was actually sent. However, if the number of possible messages $v = 1$, then the attacker has detected the transmitted message m quite unambiguously and so-called has broken the cryptogram c. To make brute force attacks more difficult, the number K of keys is chosen sufficiently large, with specific key values chosen randomly according to a uniform distribution. This is because if certain values were more likely than others, the attacker would try such keys first in his attack.

The resistance of ciphers to brute force attack is addressed by the secrecy theory. This theory describes the conditions under which a cipher is breakable and unbreakable. In this context, it should be clarified that mentioned theory only makes sense for so-called symmetric ciphers, which are cryptosystems whose both keys are secret. The counterpart of symmetric ciphers are so-called asymmetric ciphers, where the encryption key e is publicly known and only the decryption key d is secret. In this case, the indeed transmitted message can be determined unambiguously from all v possible messages. It is just the message whose encryption with the publicly known key e produces the original cryptogram c.

2. Current state

Ciphers are the subject of a science called cryptology. Its origin can be dated back to the 8th century, when the Arab scholar Al-Kindi published the first method of

Manuscript received September 5, 2022
Manuscript revised September 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.9.59

breaking cryptograms ([1], p. 17). Since then, experts have long won-dered whether there is any such thing as a cipher that cannot be broken. In fact, practical experience showed that any new cipher was broken eventually. For example, the amateur cryptologist and writer E. A. Poe wrote a short story in 1843, The Gold Bug, in which he describes the breaking of a cryptogram. And through the mouth of the protagonist, he utters the phrase "... it may well be doubted whether human ingenuity can construct an enigma of the kind which human ingenuity may not, by proper application, resolve" ([2], p. 63-64).

The proof of the existence of an unbreakable cipher was given by the American mathematician C. E. Shannon in his paper [3] in 1949. He called the mentioned type of ciphers "perfect secrecy systems". For a cipher to be unbreakable, it must satisfy conditions P1 to P3 ([4], p. 68):

- (P1): $N = C = K$, where N is the number of possible messages, C is the number of possible cryptograms, and K is the number of possible keys (and hence also the number of encryption functions).
- (P2): The key to encrypt each message is chosen randomly according to a uniform distribution.
- (P3): For each message m and each cryptogram c, there is exactly one encryption key that encrypts m into c.

In relation to condition (P1), it should be noted that in the common cryptographic literature, contrary to common perception, the term "message" also refers to a sequence of symbols that has no meaning in the language. Sequences that have meaning in a given language are distinguished in cryptography by the term "meaningful message" [5].

The best known variant of an unbreakable cipher is the Vernam cipher with a one-time key (so-called Vernam's one-time pad, e.g., [6], p. 249). A major drawback of this type of cipher is that a key of the same number of symbols as the message is required to encrypt the message, and this key must be completely random and cannot be reused.

In common practice, ciphers with shorter keys than required by perfect secrecy cryptosystems are used. To describe them, Mr. Shannon defined a so-called random cipher, for which the following applies:

- (P4): Messages are sequences of length L symbols from an alphabet consisting of B elements.
- (P5): The total number of messages $N = B^L$, whereby for the number M of meaningful messages, $M \leq N$.
- (P6): The decryption of each cryptogram c is modeled by randomly selecting a message m according to a uniform distribution from all N possible messages.

For a random cipher, the total number of all messages can be expressed as:

$$N = B^L = 2^{R \cdot L}, \tag{1}$$

where

$$R = \log_2 B \text{ [bit/symbol]} \tag{2}$$

is the entropy of one alphabet symbol. This quantity tells us that at most R bits of information can be represented by one alphabet symbol. For the number of meaningful messages, the relation is used:

$$M = 2^{r \cdot L}, \tag{3}$$

where r is the entropy of one symbol of the given language. For example, for the English language it holds ([4], p. 77) that:

$$1.0 \leq r \leq 1.5 \text{ [bit/symbol]}. \tag{4}$$

Closely related to the entropy of the language and the entropy of the alphabet is the redundancy D of the language for which:

$$D = R - r \text{ [bit/symbol]}. \tag{5}$$

The quantity D expresses how many bits of information the language symbol carries less than it theoretically could. This theoretical limit is the value of R. For example, for an English language with $B = 26$ alphabet symbols, $R = \log_2 26 = 4.7$ [bit/symbol]. For a language entropy of $r = 1.25$ [bit/symbol], the redundancy of the English language is then $D = (4.7-1.25) = 3.45$ [bit/symbol]. This means that the information capacity of the symbols in English is used to approximately 25 percent because $(1-D) = 1-3.45/4.7 = 1-0.73 \approx 0.25$. The higher redundancy of the language translates in practice to the fact that messages must be longer to encode the same amount of information. On the other hand, such messages are more robust to transmission errors.

Returning to the work [3], for a random cipher it is here given an estimate of the length L_0 of the cryptogram in which an attacker can break the cryptogram by brute force. The length

$$L_0 = \frac{\log_2 K}{D} \text{ [symbol]} \tag{6}$$

is called the "unicity distance". In this context, Mr. Shannon defined so-called "ideal secrecy systems", which are ciphers that an attacker cannot break even if the length L of the cryptograms is unlimited. When attacking a cryptogram c of this cipher, an attacker finds that more than one meaningful message may be encrypted in a given cryptogram. The most well-known variant of ciphers with ideal secrecy are ciphers that use an artificial language with zero redundancy. By relation (6), we see that indeed for redundancy $D \to 0$, the limit L_0 approaches infinity.

In addition to ciphers with perfect and ideal secrecy, Mr. Shannon also introduced so-called "practical secrecy systems". All other ciphers fall into this category, i.e. ciphers that are theoretically breakable by brute force. This type of cipher is the most widely used, and its security against brute force attack lies in the fact that the number of possible keys K is sufficiently large.

The last major result of secrecy theory is a relation for the mean number n_k of so-called "spurious keys" ([5], [7]):

$$n_k \geq 2^{H(K)-D\cdot L} - 1, \qquad (7)$$

where $H(K)$ is the key entropy, D is the language redundancy and L is the length of the cryptogram. For the commonly used method of choosing keys according to a uniform distribution, $H(K) = \log_2 K$. The quantity n_k is the mean number of keys that, when decrypted, assign a meaningful message to the cryptogram that is different from the one actually transmitted. The larger this value is, the more possible solutions the attacker will get in a brute force attack and his uncertainty about the transmitted message will be larger. He gains certainty at $n_k = 0$, where the value of L takes on the meaning of L_0 from relation (6).

3. Basic terms

In the following, we will assume that messages and cryptograms are numbers expressed in the same numerical system and have the same length. Regarding message lengths, it should be noted that in some cases the message must be extended with overhead digits before encryption. In such cases, the original message including this overhead will be considered as the message. If we use a number system with B digits and the length of both messages and cryptograms is L digits, then there are a total of $N = B^L$ numbers from the value $x = 0$ to $(N-1)$. For practical reasons, however, we will not distinguish these numbers according to the value of x, but according to their order on the numerical axis, i.e., according to the value of $s = (x + 1) \in \boldsymbol{A} = \{1, 2, 3, ..., N\}$. For the inputs i and outputs o of the encryption functions, $i, o \in \boldsymbol{A}$ and so there are in total N possible inputs and outputs, respectively. We will call the variable N the number of inputs/outputs. It should be noted here that in this paper we will consider only meaningful inputs i as messages. We will call inputs i that do not make sense in the language and context used meaningless inputs. We will also assume that the set $\boldsymbol{M}$ of all possible messages consists of a total of M messages, with $1 \leq M \leq N$.

The set of encryption respectively decryption functions consists of a total of K functions, which are given by the encryption respectively decryption key. Within this set, we will distinguish each function by an ordinal number $k \in \{1, 2, 3,..., K\}$. A particular encryption function assigns to each input $i \in \boldsymbol{A}$ a unique output $o \in \boldsymbol{A}$. Therefore, we can represent it as a permutation, i.e., as an ordered N-tuple that contains each number from the set $\boldsymbol{A}$ just once. We will write the encryption permutation given by the k-th key as $P_k = (p_1\ p_2\ ...\ p_N)$, where $p_s \in \boldsymbol{A}$ is the s-th member of the permutation. We can then express the encryption as $o = \mathrm{E}(i, k) = p_i$. For example, for $k = 2$, let us have the permutation $P_2 = (3\ 2\ 5\ 4\ 1)$. Then for $i = 3$, $p_3 = 5$ and hence $\mathrm{E}(3, 2) = 5$. The corresponding decryption function is given by the inverse permutation P_k^{-1} and then $i = \mathrm{D}(o, k) = p_o^{-1}$. For our example, $P_2^{-1} = (5\ 2\ 1\ 4\ 3)$ and hence $\mathrm{D}(5, 2) = 3$.

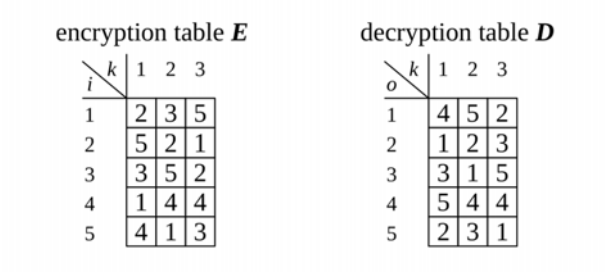
For encryption functions, the variable i is the argument and the variable o is their value. For decryption functions, the opposite is true. However, since the encryption and decryption functions form a single unit in terms of purpose, we will refer to the variable i (i.e., the input of the encryption function) as the input throughout the cryptosystem, i.e., even for the decryption function, where it acts as the function value. Similarly, we will call the variable o (i.e., the output of the encryption function) the output o in the case of decryption also, where it plays the role of an argument.

Since the number of all possible encryption or decryption functions is equal to the number of all permutations, i.e. the value $N!$, the number of all possible keys is also equal to this value. However, by (P1) we know that $K = N$ is sufficient for an unbreakable cipher, so we will assume that $1 \leq K \leq N$ for the number of keys.

If we write the individual encryption permutations in the form of columns and arrange these columns in ascending order by key number, we obtain the encryption table $\boldsymbol{E}$. With its help, for each input i and key k, we can find the encryption output $o = \boldsymbol{E}(i, k)$, where i is the row number, k is the column number, and the quantity o is the content of the table cell in the i-th row and k-th column. Similarly, from the inverse permutations, we can construct a decryption table $\boldsymbol{D}$. Using it, for each output o and key k, we can find the input $i = \boldsymbol{D}(o, k)$, where o is the row number, k is the column number, and i is the content of the table cell in the o-th row and k-th column.

To illustrate the concepts introduced above, consider a simple example where the set of inputs and outputs $\boldsymbol{A} = \{1, 2, 3, 4, 5\}$, the set of messages $\boldsymbol{M} = \{4, 5\}$ and the encryption functions are $P_1 = (2\ 5\ 3\ 1\ 4)$, $P_2 = (3\ 2\ 5\ 4\ 1)$ and $P_3 = (5\ 1\ 2\ 4\ 3)$. From the above specification, it follows that the number of inputs/outputs $N = 5$, the number of messages $M = 2$ and the number of keys $K = 3$. The inverse decryption functions are obtained by inverting the encryption permutations. It is then true that $P_1^{-1} = (4\ 1\ 3\ 5\ 2)$, $P_2^{-1} = (5\ 2\ 1\ 4\ 3)$ and $P_3^{-1} = (2\ 3\ 5\ 4\ 1)$. The encryption and decryption table of our demonstration cryptosystem is shown in Fig. 2.

From the encryption table (shown on the left) we can easily determine that, for example, for input $i = 1$ and key number $k = 3$, the output $o = \boldsymbol{E}(i, k) = \boldsymbol{E}(1, 3) = 5$. The figure on the right shows the corresponding decryption table $\boldsymbol{D}$. Using it, we can easily find out what the input was in the above encryption. We know that the output $o = 5$ and the key has the number $k = 3$. Then the output of the decryption, i.e., the input during encryption $i = \boldsymbol{D}(o, k) = \boldsymbol{D}(5, 3) = 1$, which is indeed the original message.



i \ k	1	2	3
1	2	3	5
2	5	2	1
3	3	5	2
4	1	4	4
5	4	1	3

o \ k	1	2	3
1	4	5	2
2	1	2	3
3	3	1	5
4	5	4	4
5	2	3	1

Fig. 2: Example of an encryption and decryption table

4. Model

From the point of view of the resistance of ciphers to brute force attack, we are mainly interested in the decryption table ***D***. In our example in Figure 2, we see that if an attacker intercepts, say, a cryptogram $c = o = 1$ in the channel, he can determine from the corresponding (i.e., first) row of the decryption table ***D*** what the possible inputs of the cipher were. For keys $k = 1$, resp. 2, resp. 3 these inputs could be $i = 4$, resp. 5, resp. 2. He can exclude the input $i = 2$, since it is a meaningless input. Thus, the attacker concludes that either message $m = 4$ or message $m = 5$ is encrypted in cryptogram $c = 1$. In our cryptosystem, these are all possible messages, so the attacker has gained nothing by his attack. He already knew that one of all possible messages was being transmitted when he eavesdropped on the cryptogram. However, if the attacker intercepted the cryptogram $c = 3$, then by analogy he would find that the message $m = 5$ is encrypted in it. This conclusion of his is quite unambiguous, since the other possible inputs (i.e., $i = 1$ and 3) are not messages.

The above example shows that the decryption table should be constructed in such a way that as many different messages as possible can be found in each of its rows. Let us now look at this requirement more generally. Let us call the number of distinct messages in each row of the decryption table the valence of the corresponding output. Formally, we will define the valence v_o of the o-th output as the number of distinct messages obtained by decrypting that output with all possible keys. The minimum possible value of valence $v_o = 0$. In this case, for all keys, a given output is an image of only meaningless inputs, so such an output cannot appear in the transmission channel. We will therefore call it an absurd output. Another outputs are, for at least one key, the image of some message, i.e. their valence is at least 1. We will call such outputs cryptograms. In terms of the maximum possible value of the valence of any output, this value obviously cannot be larger than the total number of M messages, and also cannot be larger than the total number of K columns of the decryption table. Thus, we can write that:

$$0 \leq v_0 \leq \min\{M, K\}. \quad (8)$$

In our example, $v_1 = 2$, $v_2 = 0$, $v_3 = 1$, $v_4 = 2$ (there are three messages in line 4, but one is there twice) and $v_5 = 0$. Thus the output $o = 2$ and 5 is an absurd output (i.e., it will never be transmitted in the transmission channel) and the other outputs are cryptograms. Cryptogram $c = 3$ has a valence equal to one and thus by brute force attack the message transmitted in it is uniquely detectable. On the other hand, cryptograms $c = 1$ and 4 are so-called unbreakable, since any of the M possible messages may be encrypted in them. In cryptography, a pessimistic viewpoint is used to assess security, so we will evaluate the security of the entire cipher according to the worst case, i.e., according to the cryptogram that has the smallest valence of all. We will call that parameter the minimum valence V of the cipher. For a range of values of this quantity, the following is of course true:

$$1 \leq V \leq \min\{M, K\}. \quad (9)$$

Our illustrative cipher has a minimum valence $V = \min\{2, 1, 2\} = 1$. In this context, note that in addition to the decryption table, the valence V also depends on the message set. For the same decryption table, when we change the message set to $M = \{1, 3\}$, then the outputs $o = 1$ and 4 are absurd, and the valence of the other outputs implies that $V = \min\{2, 2, 2\} = 2$, which is a higher value compared to the original example. Thus, it can be concluded that the resistance to brute force attack depends not only on the cipher itself but also on the message language.

To maximize the value V, the decryption table should be constructed in such a way that each row of the table contains as many different messages as possible. To do this, it is advisable, among other things, that the messages in the rows of the table are not repeated. This leads to the requirement that the rows of the decryption table be so-called K-permutations, which are ordered K-tuples of elements from all N possible elements, where each element can occur at most once in a given K-tuple. If we recall that the columns of the table are permutations, the decryption table should take the form of a so-called Latin rectangle $N \times K$, where $K \leq N$. For a Latin rectangle, it is true that in each row and in each column any element $p_s \in \boldsymbol{A}$ occurs at most once (e.g., [8], p. 385). We will call a cipher with N inputs and outputs, M messages and K keys, whose decryption table consists of a Latin rectangle, a Latin cipher with parameters (N, M, K).

The fact whether modern ciphers can be described by a Latin rectangle is not completely obvious from the point of view of K-permutations, so we discuss it now. Stream ciphers are purposefully constructed so that changing the key changes the encryption sequence of the pseudorandom generator. Therefore, the same message with a different key will be encrypted into a different cryptogram each time, so stream ciphers belong in the category of Latin ciphers. In the case of modern block ciphers, the message block is first

merged with the key by a suitable mathematical operation *f* (usually XOR). For example, in the AES cipher, this is the initial operation AddRoundKey (e.g., [9], p. 15). Other subsequent operations are constructed such that for a given key, each possible message block is assigned a unique block of the cryptogram. However, as a consequence of the merge operation *f*, the uniqueness of the assignment holds even in the situation where the message block is the same and the key changes. It then follows that modern block ciphers also belong to the category of Latin ciphers.

Because of the significant size of the decryption tables of modern ciphers and also because of the random occurrence of messages in the input set, the valence *V* of these ciphers cannot be determined accurately at present. Here we exploit the fact that in the case of the Latin cipher, the rows of the decryption table are *K*-permutations of *N* elements (i.e., inputs), of which *M* elements have this property of being messages. In fact, if we look at the decryption table not in terms of the numerical values of the inputs, but in terms of whether or not a given input is a message, then we conclude that the rows of the table become independent of each other and that the occurrence of messages in the rows of the table follows a hypergeometric distribution.

The first observation follows from the fact that messages occur randomly in column permutations. And since the column permutations of the cipher are different and independent of each other, then the resulting occurrence of messages in the rows is random and independent of the occurrence of messages in the other rows.

The second observation, the fact that the occurrence of messages in rows follows a hypergeometric distribution, follows from the very definition of the type of distribution mentioned. Namely, a hypergeometric distribution describes the process of selecting *K* elements randomly from a set of *N* elements without returning, where *M* elements out of all *N* elements have a certain property that the remaining (*N*−*M*) elements do not have (e.g., [10], p. 60). In our case, the *N* elements are the possible inputs from which *K* inputs are randomly selected for each row of the decryption table and *M* is the number of messages. The distinguishing property of the elements here is whether or not a given input is a message. For the probability $P(X = v)$ of a hypergeometric distribution, the following holds:

$$P(X = v) = \frac{\binom{M}{v} \cdot \binom{N-M}{K-v}}{\binom{N}{K}}, \qquad (10)$$

where the variable *v* is the number of messages in a row of the decryption table (i.e. the valence of the corresponding output), *M* is the total number of messages, *K* is the number of keys, and *N* is the number of inputs/outputs. The values of the variable *v* that have a non-zero probability of occurrence are in the interval:

$$\max\{0, M + K - N\} \leq v \leq \min\{M, K\}. \qquad (11)$$

The two boundaries of the mentioned interval are plotted in Figure 3 and Figure 4, respectively, for the example of ciphers with the number of inputs/outputs $N = 50$. Figure 3 shows the lower bound $Q = \max\{0, M+K-N\}$ and Figure 4 shows the upper bound $U = \min\{M, K\}$. The two bounds take the form of parts of two divergent planes.

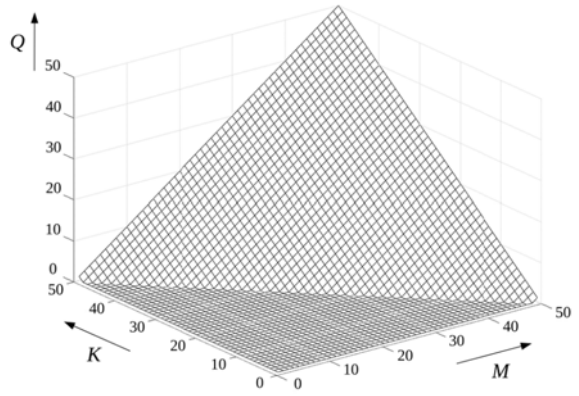


Fig. 3: Lower bound *Q* of the valence of outputs for ciphers with $N = 50$

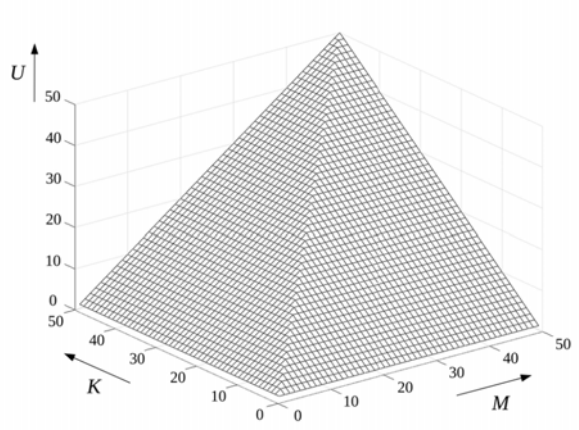


Fig. 4: Upper bound *U* of valence of outputs for ciphers with $N = 50$

We refer to Figure 5 to illustrate the hypergeometric distribution. In the upper part of the figure we see the distribution for $N = 10$, $K = M = N/2 = 5$, for which the valence *v* is in the range 0 to 5. The bottom of the figure then shows the distribution for $N = 100$, $K = M = N/2 = 50$, where the valence *v* of the outputs is in the range 0 to 50. Note that for large values of *N*, the probability of many valence values is close to zero. For example, for the lower graph, for the probabilities of the two extreme values of valence *v*, $P(v = 0) = P(v = 50) = 9{,}9 \cdot 10^{-30}$. On the other hand, for the mean valence value $E = 25$, it holds that $P(v = 25) = 0.158$. Thus, cryptograms with valence close to the mean value occur in large-scale cryptosystems with

probabilities that are orders of magnitude higher than the probabilities of occurrence of other outputs.

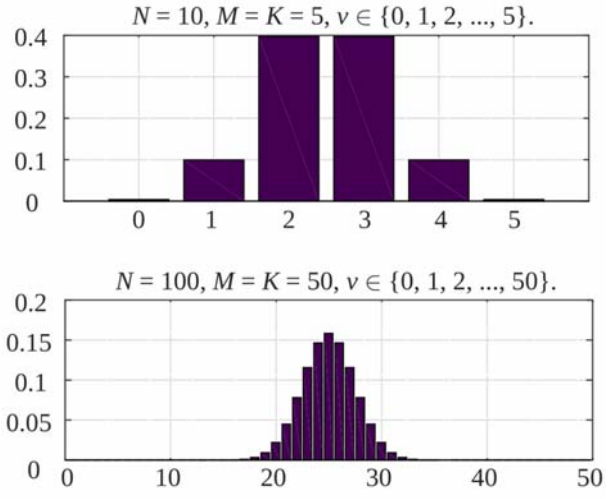


Fig. 5: Examples of hypergeometric distributions

At the end of the characterization of the hypergeometric distribution, we will state the relation for the mean value E of the random variable X according to this distribution:

$$E = \frac{M \cdot K}{N}. \tag{12}$$

We can physically interpret the mean value E of the valence of outputs as the Latin cipher assigns to each message m a total of K different outputs out of a total of N outputs. Then, on average, each of the N outputs is assigned a total of $E = M{\cdot}K/N$ messages. An example of the dependence of the mean value E of the valence of outputs on the number of keys K and the number of messages M for ciphers with $N = 50$ inputs/outputs is shown in three-dimensional form in Fig. 6.

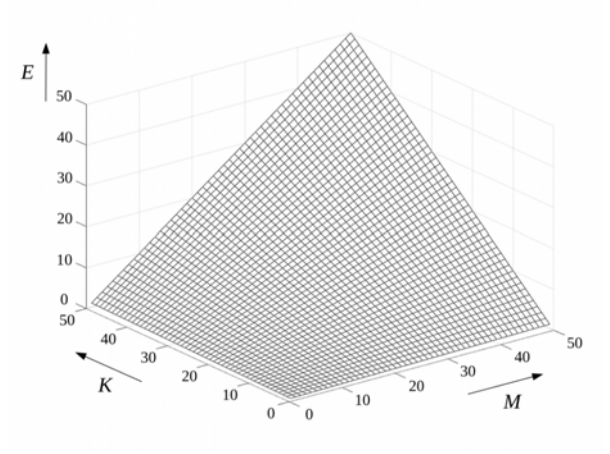


Fig. 6: Dependence of the mean value E for ciphers with $N = 50$

Then in Fig. 7, for the same cipher, we have the dependence of the lower bound Q, the upper bound U and the mean value E for the situation where $M = K$ to compare.

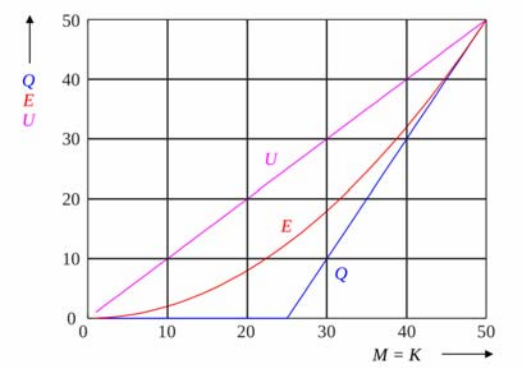


Fig. 7: Lower bound Q, upper bound U and mean E of the valence of outputs for ciphers with $N = 50$

We again use a pessimistic approach to assess the security of a cipher against a brute force attack, calling the corresponding parameter the lower bound W of the valence of ciphers. In relation (11), the pessimistic view is represented by the left inequality corresponding to the quantity Q. Zero valence only holds for absurd outputs and so the lower bound W for the valence of cryptograms will be:

$$W = \max\{1, M + K - N\}. \tag{13}$$

We will now modify this relationship into a function:

$$W = \begin{cases} 1, & \text{if } (M + K) \leq (N + 1), \\ M + K - N, & \text{if } (M + K) > (N + 1), \end{cases} \tag{14}$$

recalling that $1 \leq M \leq N$ and $1 \leq K \leq N$.

The probability $P(X = W)$ is always non-zero and so in the decryption tables of some ciphers there must be rows that correspond to cryptograms with valence W. And since this is a non-zero lower bound, the value of W will also be the valence V of the corresponding cipher. Ciphers with decryption tables that do not have such rows will naturally have a higher valence V. However, from a pessimistic point of view, any Latin cipher with N inputs/outputs, M messages, and K keys is guaranteed to have at worst a valence $V = W$ according to relation (14).

5. Discussion

The lower bound W of the valence of ciphers is the main contribution of this paper, so we now discuss it in more detail. In Figure 8, we have a three-dimensional representation of the values of W as a function of the number of keys K and the number of messages M for ciphers with $N = 50$ inputs/outputs. From the above graph it can be seen that the points of the lower bound W of the valence lie in two planes. The horizontal plane includes ciphers with $W = 1$.

These are all cryptosystems for which $(M+K) \leq (N+1)$ according to (14). For other ciphers, it holds that $(M+K) > (N+1)$. In this case, the values of W lie in the skew plane given by the equation $W = (M+K-N)$.

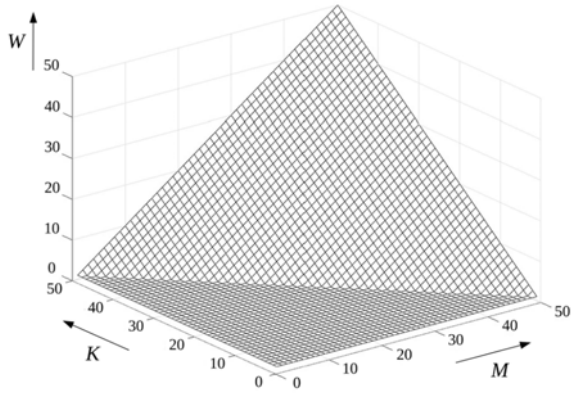


Fig. 8: Lower bound W of the valence of ciphers with $N = 50$ as a function of the number of keys K and the number of messages M

Fig. 9 shows a similar graph for $N = 10$. This graph is clearer and we therefore discuss the results on it. The square points represent ciphers for which the number of keys is equal to the number of inputs/outputs, i.e., $K = N = 10$. For these ciphers, we can use function (14) to derive that $W = M$, i.e., an attacker can use the brute force method to determine that any one of all M possible messages may be encrypted in any given cryptogram. These are thus unbreakable ciphers, which Mr. Shannon called perfect secrecy systems. Condition $N = K$ is inconsistent with condition (P1), where $N = C = K$. However, this inconsistency is only apparent, since the equality $N = C$ in existing secrecy theory expresses the requirement that the number of inputs and the number of outputs of the encryption function be equal. For Latin ciphers, this equality is given by their definition and the requirement $N = K$ is therefore sufficient. The equality $N = K$ implies that the length of the keys must be equal to the length of the inputs/outputs. In addition, the keys must be random and unique for each message, so these ciphers, while completely immune to brute force attack, are very rarely used.

For ciphers represented by circular points, $2 \leq W \leq (M-1)$. Using the brute force method, an attacker finds that in a given cryptogram, any of the W possible messages may be encrypted. He does not know which of them was actually transmitted, but on the other hand he knows safely that it was not transmitted some of the remaining $(M-W)$ messages. These partially unbreakable ciphers have been named by Mr. Shannon as systems of ideal secrecy. We can see from the figure that they require either the use of a large number of keys or the removal of the greatest possible amount of redundancy from the language used. Alternatively, the two can be combined.

The extremum of the method based on maximizing the number of keys is represented by the points on the strong dashed line, where $K = N-1 = 9$. The function (14) then implies that the valence of these ciphers is $W = M-1$. Unfortunately, a high number of keys is the same problem that perfect secrecy systems have.

The extremes of the redundancy elimination based method are represented by points on a strong continuous line. By removing all redundancy, every possible input becomes a message, i.e., $M = N = 10$. Then, according to (14), the valence $W = K$. Unfortunately, the redundancy elimination method is not yet significantly applicable, as there are currently no sufficiently powerful and fast compression algorithms available for natural languages.

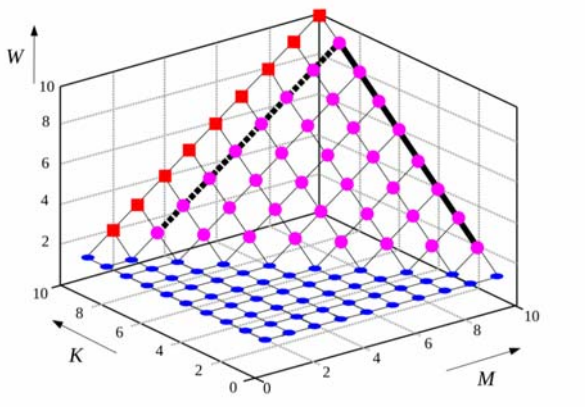


Fig. 9: Lower bound W of the valence of ciphers with $N = 10$

The oval points represent ciphers for which $W = 1$. This type of cipher is thus theoretically breakable, and therefore Mr. Shannon called them practical secrecy systems. They are the most widespread, and their security lies in a sufficiently large number of keys. So large that in the time T of the cipher's resistance, i.e., the time for which the messages are supposed to remain secret, not all the keys can be tested. Currently, it is generally recommended that $K \geq 2^{128}$ (e.g., [11], pp. 59 and 53).

We now show that the model presented above allows us to derive all the relevant results of existing secrecy theory. We start with relation (7) for the mean value of the false keys n_k. The value of n_k plus the correct key, i.e., (n_k+1), is effectively the mean number of keys that assign a meaningful message to the output. By definition, this value should correspond to the mean value of the valence E according to relation (12). From the derivation below, where we have used the substitutions in (1), (3) and (5), it is clear that this is indeed the case.

$$n_k + 1 \geq 2^{H(K)-D\cdot L} = \frac{2^{H(K)}}{2^{D\cdot L}} = \frac{K}{2^{(R-r)\cdot L}} = \\ = \frac{K \cdot 2^{r\cdot L}}{2^{R\cdot L}} = \frac{K \cdot M}{N} = E\,. \quad (15)$$

Similarly, it can be shown that by using the mean valence value, relation (6) can be derived to determine the unicity distance L_0. The proof is given in the appendix at the end of the paper. The condition that is used here is that to unambiguously decipher a cryptogram, its valence must be $v = 1$. If we relate this condition to the mean valence value, we obtain the relation (6) just mentioned.

In this context, we can additionally define a pessimistic unicity distance L_0 using a lower bound on the minimum valence of ciphers. As we already know, a brute force attack will lead to an unambiguous result for cryptograms whose valence $v = 1$. Thus, by the lower bound in (14), it must hold:

$$M + K = N + 1, \quad (16)$$

and after inserting relations (1) and (3) we obtain the equation:

$$2^{r\cdot L_0} + K = 2^{R\cdot L_0} + 1. \quad (17)$$

The value of L_0, which is the solution to the above equation, tells us that for Latin ciphers (N, M, K) there are cryptograms from length L_0 onwards that are already breakable. For example, for a simple substitution cipher where $K = 26!$ and for an English text with $R = \log_2 26 = 4.7$ [bit/symbol] and $r = 1.5$ [bit/symbol], the solution to Equation (17) is $L_0 \approx 19$ symbols. According to relation (6), the analogous value comes out to 26 symbols. However, there is no contradiction here. The value according to (17) tells us that in some ciphers (N, M, K) there exist cryptograms that are breakable from a length of 19 symbols. And the value according to (6) tells us that cryptograms are, on average, breakable from a length of 26 symbols. So the first value is pessimistic and the second is average.

If we want to find the mean valence of the cryptograms instead of the mean valence of the outputs, we have to exclude the absurd outputs from relation (12). Since it is clear that there are $N{\cdot}P(X{=}0)$ absurd outputs in N outputs, the total number of cryptograms (i.e., outputs with valence $v > 0$) is then equal to $C = N{\cdot}[1{-}P(X{=}0)]$. Using this relation, we can then define the sought-after mean valence of the cryptograms Z:

$$Z = \frac{M \cdot K}{C} = \frac{M \cdot K}{N \cdot [1 - P(X = 0)]}\,. \quad (18)$$

Returning to the lower bound W of the valence of ciphers, it has already been mentioned that for larger values of N the probabilities of cryptograms with extreme valence values are often close to zero. For example, in Figure 10 we have a logarithmic plot of the probability p_W, which is the probability of occurrence of cryptograms with valence W, for ciphers with $N = 100$ and $K = 50$ versus the number of messages M. We see here that for a magnitude M close to $N/2$ the mentioned probability is very small. In particular, for example, for $M = 51$, $p_W \approx 5{,}1{\cdot}10^{-28}$.

It is clear from the figure that for larger values of N the lower bound on W is quite pessimistic. Then, a parameter that we call the statistical estimate S of the lower bound of the valence of ciphers can be useful. The aforementioned parameter is based on the fact that the lower bound of the valence is increased by the valence of cryptograms whose overall probability of occurrence is negligible for a given scenario. Let us introduce the quantity p_S, which is the probability of occurrence of cryptograms with valence less than S, i.e., with valence in the range of values W to (S−1).

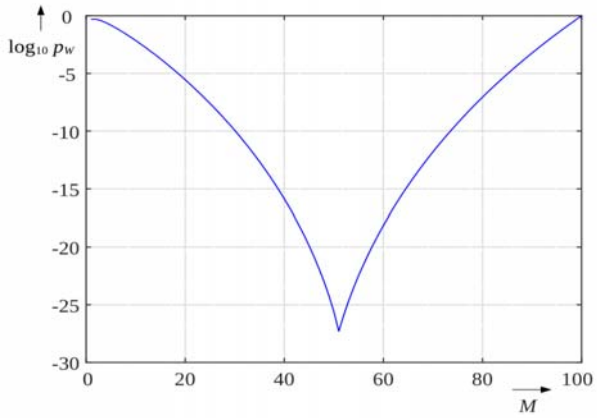


Fig. 10: Dependence of the probability p_W of occurrence of cryptograms with valence W on the number of messages M for ciphers with $N = 100$ and $K = 50$

Let us now have a set of n ciphers in total, each containing N outputs. In such a set, there are $n{\cdot}N$ outputs, of which there will be on average $p_S{\cdot}n{\cdot}N$ such cryptograms whose valence v is in the range of W to (S−1). Let us stipulate that the number of these cryptograms should be at most 1 on average, i.e., in the set of all n ciphers, there will be on average at most one such cryptogram whose valence is in the range of W to (S−1). Formally, then:

$$p_S \cdot n \cdot N \leq 1\,. \quad (19)$$

Then, on average, there will be at most one cipher in our set with valence V in the range W to (S−1), and the remaining (n−1) ciphers will have minimum valence $V \geq S$. We call this valence the statistical estimate S of the lower bound on the valence of ciphers with estimation error $\delta = 1/n$. In practice, we find the quantity S by finding the largest such integer $S \in \langle W, U\rangle$ for the specified parameters N and δ such that the probability p_S from equation (20) below satisfies condition (19). If no such value exists, then $S = W$.

$$p_S = \sum_{v=W}^{S-1} P(X = v). \quad (20)$$

We compare the dependence of the statistical estimate S of the lower bound of the valence and the dependence of the lower bound W of the valence of the ciphers in Figure 11. Both dependencies hold for ciphers with $N = 100$ inputs, resp. outputs, where the number of messages is the same as the number of keys, i.e., $M = K$. For comparison, the dependence of the upper bound U and the mean value Z of the valence of the cryptograms in the mentioned ciphers is also shown. The dependence of the quantity S confirms the fact that the probability of cryptograms with small valences is very low for those arguments M and K that are close to the value $N/2 = 50$. For example, for $M = K = 50$, the value of $W = 1$, but the value of $S = 11$. This can be explained by the fact that the overall probability of occurrence of cryptograms with valence from 1 to 10 in our case is equal to $p_S = 1{,}1 \cdot 10^{-9}$. It follows from $\delta = 1/n = 10^{-6}$ that $n = 10^6$ and so, after substituting in $p_S \cdot n \cdot N = 1{,}1 \cdot 10^{-9} \cdot 10^6 \cdot 100 = 0{,}11$, we see that inequality (18) is satisfied. We can interpret the value of $S = 11$ to mean that if we randomly select 10^6 Latin ciphers with parameters $N = 100$, $M = 50$, and $K = 50$, then on average for $(10^6 - 1)$ of these ciphers the minimum valence V will be at least 11. And on average, only one cipher will have a minimum valence that is less than 11.

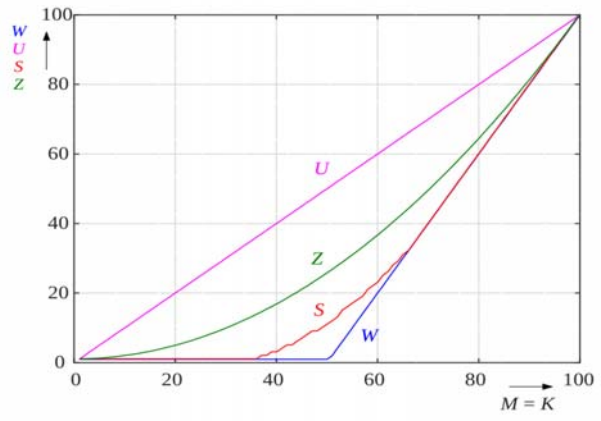


Fig. 11: Dependence of the statistical estimate S of the lower bound of the valence of the ciphers with estimation error $\delta = 10^{-6}$. The dependence applies to ciphers with $N = 100$, whereby $M = K$

6. Conclusion

Overall, it can be stated that the paper extends the existing Shannon's theory of secrecy systems. The extension consists in replacing the random cipher by a more adequate Latin cipher and in introducing the notion of valence of a cryptogram. In the paper, a Latin cipher (M, K, N) is a cipher with K keys, N outputs and N inputs, of which M inputs make sense in a given language and context, i.e. there are M messages. The defining feature of a Latin cipher is the property that each of its inputs, when encrypted with all possible K keys, takes the form of K mutually distinct outputs. Inversely, each output, when decrypted with all possible K keys, takes the form of K mutually distinct inputs.

In this paper, it is shown that all modern ciphers are Latin ciphers and so the number of messages v produced by decrypting their arbitrary output with all possible keys can be modeled by a hypergeometric distribution according to relation (10). Decrypting a cryptogram with all possible keys is called a brute force attack and the quantity v is called the valence of the output. If $v = 1$, the attacker detects the transmitted message from the intercepted cryptogram uniquely (so-called a breakable cryptogram). If $v = M$, then the cryptogram is unbreakable because any of the M possible messages could have been transmitted in it. And if $1 < v < M$, then any of the v possible messages could have been transmitted in the cryptogram and the attacker does not know which one. On the other hand, the attacker knows that none of the remaining $(M-v)$ messages were transmitted. We have called such cryptograms partially unbreakable.

The mean value E of the valence of the outputs, which is given by relation (12), provides an approximate assessment of the resistance of cryptograms to brute force attack. If $E = M$, then every cryptogram of the cipher is unbreakable, and if $E \leq 1$, then on average every randomly chosen cryptogram is breakable. A higher cipher security guarantee is provided by the lower bound W on the valence of ciphers according to relation (14). This quantity tells us that every cryptogram of an arbitrary cipher (M, K, N) has valence $v \geq W$. Using the quantity W, we can thus classify not only individual cryptograms but entire ciphers into breakable, partially unbreakable, and unbreakable.

It is further shown in the paper that the proposed model allows to derive all previously known insights of the theory of secrecy systems, which are perfect or ideal or practical secrecy systems, unicity distance and number of spurious keys. The model is thus also suitable for pedagogical purposes, as an introduction to the theory of secrecy systems.

References

[1] Singh S.: The Code Book: The Secret History of Codes and Code-Breaking. Fourth Estate, London 2000.

[2] Poe E. A.: The Gold-Bug. D. Estes & company, Boston 1899.

[3] Shannon, C. E.: Communication Theory of Secrecy Systems. Bell System Technical Journal. Vol. 28, Issue 4, Oct. 1949, pp. 656–715.

[4] Stinson D. R., Paterson M. B.: Cryptography. Theory and Practice. CRC Press, Boca Raton 2019.

[5] Hellman M.: An extension of the Shannon theory approach to cryptography. In IEEE Transactions on

Information Theory, vol. 23, no. 3, pp. 289-294, May 1977.
[6] Hoffstein J., Pipher J., Silverman J. H.: An Introduction to Mathematical Cryptography. Springer, N. York 2008.
[7] Beauchemin P., Brassard G.: A generalization of Hellman's extension to Shannon's approach to cryptography. In Journal of Cryptology, vol. 1, no. 3, pp. 129–131, Oct 1989.
[8] Brualdi, R. A.: Introductory Combinatorics. Pearson, London 2009.
[9] Dworkin M. J. et al.: Advanced Encryption Standard (AES). NIST FIPS - 197. National Institute of Standards and Technology, Gaithersburg 2001.
[10] Montgomery D. C., Runger G. C.: Applied Statistics and Probability for Engineers. Wiley, Hoboken 2018.
[11] Barker E.: Recommendation for Key Management: Part 1 – General. NIST SP 800-57, Part 1, Rev. 5. National Institute of Standards and Technology, Gaithersburg 2020.

Appendix

The aim of the appendix is to prove that condition:

$$E = \frac{M \cdot K}{N} = 1 \qquad \text{(A1)}$$

leads to equation (6):

$$L_0 = \frac{\log_2 K}{D}. \qquad \text{(A2)}$$

First, in relation (12) for the mean value E, we make substitutions according to (1) and (3). We obtain:

$$E = \frac{M \cdot K}{N} = \frac{2^{r \cdot L} \cdot K}{2^{R \cdot L}}. \qquad \text{(A3)}$$

After adjusting and inserting (5) we have:

$$E = \frac{K}{2^{(R-r) \cdot L}} = \frac{K}{2^{D \cdot L}}. \qquad \text{(A4)}$$

For the unicity distance L_0, (A1) must hold, so:

$$E = \frac{K}{2^{D \cdot L_0}} = 1. \qquad \text{(A5)}$$

After logarithmizing and modifying the above equation, we finally obtain equation (A2), which should have been proved.

Karel Burda received the M.S. and Ph.D. degrees in Electrical Engineering from the Liptovsky Mikulas Military Academy in 1981 and 1988, respectively. During 1988-2004, he was a lecturer in two military academies. At present, he works at Brno University of Technology. His current research interests include the security of information systems and cryptology.