Unconditionally secure ciphers with a short key for a source with unknown statistics

Boris Ryabko

Federal Research Center for Information and Computational Technologies
Novosibirsk State University

Abstract

We consider the problem of constructing an unconditionally secure cipher with a short key for the case where the probability distribution of encrypted messages is unknown. Note that unconditional security means that an adversary with no computational constraints can obtain only a negligible amount of information ("leakage") about an encrypted message (without knowing the key).

Here we consider the case of a priori (partially) unknown message source statistics. More specifically, the message source probability distribution belongs to a given family of distributions. We propose an unconditionally secure cipher for this case. As an example, one can consider constructing a single cipher for texts written in any of the languages of the European Union. That is, the message to be encrypted could be written in any of these languages.

Keywords: *cryptography, unconditionally secure cipher, entropically-secure symmetric encryption scheme, indistinguishability, data compression, universal code.*

1 Introduction

The concept of unconditional security is very attractive to cryptography and has found many applications since C. Shannon described it in his famous article [1]. The concept refers to secret-key cryptography involving three

participants Alice, Bob and Eve, where Alice wants to send a message to Bob in secret from Eve, who has the ability to read all correspondence between Alice and Bob. To do this, Alice and Bob use a cipher with a secret key k (i.e. a word from some alphabet), which is known to them in advance (but not to Eve). When Alice wants to send some message m, she first encrypts m using key k and sends it to Bob, who in turn decrypts the received encrypted message using the key k. Eve also receives the encrypted message and tries to decrypt it without knowing the key. The system is called unconditionally secure, or perfect, if Eve, with computers and other equipment of unlimited power and unlimited time, cannot obtain any information about the encrypted message. Not only did C. Shannon provide a formal definition of perfect (or unconditional) secrecy, but he also showed that the so-called one-time pad (or Vernam cipher) is such a system. One of the specific properties of this system is the equivalence of the length of the secret key and the message (or its entropy). Moreover, C. Shannon proved that this property must be true for any perfect system. Quite often this property has limited practical application as many modern telecommunication systems forward and store megabytes of information and the requirement to have secret keys of the same length seems to be quite stringent. There are, therefore, many different approaches to overcoming this obstacle. These include the ideal systems proposed by C. Shannon [1], the so-called honeycomb cipher proposed by Jewels and Ristenpart [2], the so-called entropy security proposed by Russell and Wang [3] and some others developed in recent decades [4–11].

The present work is concerned with entropically secure ciphers.

It is important to note that an entropically secure cipher is not perfect, and Eve may obtain some information about the message — the property referred to as "leakage," see the definition below, but this leakage can be made negligible. On the other hand, an entropically secure cipher makes it possible to significantly reduce the key length (compared to the perfect cipher).

Recently, an entropically secure cipher has been proposed for the case where encrypted messages have a known distribution, and for the case where messages are generated by a Markov chain [11]. In the case of a known distribution, the length of the secret key is independent of message length, while in the case of a Markov chain, the length of the key grows logarithmically with message length; in both cases the length of the key depends on the amount of leakage.

In this paper we consider the situation where encrypted messages obey an unknown (or partially unknown) probability distribution. We propose an entropically secure cipher for which the key length depends on universal code (or data compressor) used for encoding the source and on the admissible leakage of the cipher. In a sense, the problem under consideration includes as special cases the previously solved problems with known probability distribution and the case where messages are generated by a Markov chain.

The construction of the cipher is based on entropically secure ciphers [3, 5, 10, 11] and universal coding [12]. It is worth noting that the proposed cipher uses data compression and randomisation, both of which are quite popular in unconditional security, cf. [13–15] and [15, 16], respectively.

2 Definitions and preliminaries

2.1 Basic concepts

We consider the problem of symmetric encryption, where Alice wants to securely transmit a message to Bob. The messages are n-letter binary words, they obey a certain probability distribution p defined on the set $\{0,1\}^n, n \geq 1$. This distribution is only partially known, i.e. it is known that p belongs to some given set P, $P \subset R^n$. Alice and Bob have a shared secret key $K = K_1 ... K_k$, and Alice encrypts the message $M \in \{0,1\}^n$ using K and possibly some random bits. Then she sends the word $cipher(M,K)$ to Bob, who decrypts the received $cipher(M,K)$ and obtains M. The third participant is a computationally unconstrained adversary Eve, who knows $cipher(M,K)$ and distribution p, and wants to find some information about M without knowing K.

Russell and Wang [3] suggested a definition of entropic security which was generalised by Dodis and Smith [5] as follows: A probabilistic map Y is said to hide all functions on $\{0,1\}^n$ with leakage ϵ if, for every adversary A, there exists some adversary $\hat{A}$ (who does not know $Y(M)$) such that for all functions f,

$$| Pr\{A(Y(M)) = f(M)\} - Pr\{\hat{A}(\,) = f(M)\} | \leq \epsilon. \qquad (1)$$

(note that $\hat{A}$ does not know $Y(M)$ and, in fact, she guesses the meaning of the function $f(M)$.) In what follows, the probabilistic map Y will be $cipher(M,K)$ and f is a map $f : \{0,1\}^n \to \{0,1\}^*$.

Definition 1. The map $Y()$ is called ϵ-entropically secure for family probability distributions P if $Y()$ hides all functions on $\{0,1\}^n$ with leakage of ϵ, whenever $p \in P$.

Note that, in a sense, Definition 1 is a generalisation of Shannon's notion of perfect security. Namely, if we take $\epsilon = 0$ and $Y = cipher(M, K)$ and $f(x) = x$, we obtain that for any M

$$|Pr\{A(cipher(M,K)) = M\} - Pr\{\hat{A}(\ ) = M\}\ | \ = 0$$

So, A and $\hat{A}$ obtained the same result, but A estimates the probability based on $cipher(M, K)$, whereas $\hat{A}$ does it without knowledge of $cipher(M, K)$. Thus, the entropic security (1) can be considered as a generalisation of the Shannon's perfect secrecy.

We will use another important concept, the notion of indistinguishability.

Definition 2 A randomised map $Y : \{0,1\}^n \to \{0,1\}^n, n \geq 1$, is ϵ-indistinguishable for some family of destributions $\mathbf{P}$ and $\epsilon > 0$ if there is a probability distribution G on $\{0,1\}^n$ such that for every probability distribution $p \in \mathbf{P}$ we have

$$SD(Y(M), G) \leq \epsilon,$$

where for two distributions A, B

$$SD(A,B) = \frac{1}{2} \sum_{U \in \mathbf{M}} |Pr\{A = U\} - Pr\{B = U\}| \, .$$

Importantly, G is independent of $Y(M)$.

Dodis and Smith [5] showed that the concepts of ϵ-entropic security and ϵ-indistinguishability are equivalent up to small parameter changes.

2.2 ϵ-entropically secure ciphers for distributions with bounded min-entropy

In 2006 [3], the first entropy secure cipher was developed for probability distributions with a limited value of the so-called minimum entropy, which is defined as follows

$$h_{min}(p) = -\log \max_{a \in A} p(a) \, . \tag{2}$$

where p is a probability distribution, $\log = \log_2$. The Russell and Wang [3] cipher was generalized and developed by Dodis and Smith [5] and their result can be formulated as follows:

Theorem 1 [5]. *Let p be a probability distribution on $\{0,1\}^n, n > 0$, whose min-entropy is not less then $h, h \in [0,n]$. Then there exists an ϵ-entropically secure cipher with the k-bit key where*

$$k = n - h + 2log(1/\epsilon) + 2. \tag{3}$$

Let's denote this cipher as $cipher_{rw-ds}$.

In a sense, this cipher generalizes the perfect Shannon cipher as follows: In a perfect cipher the key is the word from $\{0,1\}^n$, while in an entropically secure cipher the key belongs to the 2^k-element subset $K \subset \{0,1\}^n$, which is a so-called small-biased set. Informally, this means that for any $m \leq n$ and a uniformly chosen binary word $u \in \{0,1\}^m$, for any m positions $i_1 i_2, ..., i_m$, the probability that $K_{i_1}, K_{i_2}, ..., K_{i_m} = u$ is close to 2^{-m}. (This construction is based on some deep results in combinatorics [5,17,18].) Thus, the key length decreases from n to k. Note that the leakage ϵ and hence the summand $2\log(1/\epsilon) + 2$ depends on the size of the "small-biased set" 2^k (In general, larger k implies smaller ϵ.)

2.3 ϵ-entropically secure ciphers with reduced secret key

In equality (3), the linearly increasing summand $n - h$ depends on the min-entropy h. So, it seems natural to transform the set $\{0,1\}^n$ so as to reduce the min-entropy of the original distribution p and hence the summand $n - h$. In [11] this approach was realised as follows: let there be a set of probability distributios $\mathbf{P}$ defined on $\{0,1\}^n, n \geq 1$. The key part of the cipher is such a randomised map $\phi : \{0,1\}^n \to \{0,1\}^{n^*}, n^* \geq n$, that there exists a map ϕ^{-1} (i.e $\forall\, u \;\; \phi^{-1}(\phi(u)) = u$) and a min-entropy of the transform probabiity distribution π_p is close to n^* (here the distribution π_p is such that $p(u) = \sum_{v:\phi^{-1}(v)=u} \pi_p(v)$). And then the $cipher_{rw-ds}$ can be applied to $\phi(m)$ with a shorter key, because the difference $n^* - h_{min}(\pi_p)$ will be less that $n - h_{min}(p)$, see (3). Thus, the smaller $\sup_{p\in P}(n^* - h_{min}(\pi_p))$, the shorter the secret key. The described cipher is based on data compression and randomisation and denoted in [11] by $cipher_{c\&r}$. The following theorem describes its properties.

Theorem 2 [11]. *Suppose there is a family P of probability distributions defined on $\{0,1\}^n$ and there is a randomised mapping $\phi : \{0,1\}^n \to \{0,1\}^{n^*}, n^* \geq n$ for which there exists a mapping ϕ^{-1} and let*

$$\sup_{p\in P}(n^* - h_{min}(\pi_p)) \leq \Delta\,. \tag{4}$$

for some Δ*. Then*

i) $cipher_{c\&r}$ *is* ϵ*-entropically secure with secret key length* $\Delta + 2\log(1/\epsilon) + 2$*, and*

ii) $cipher_{c\&r}$ *is* ϵ*-indistinguishable with secret key length* $\Delta + 2\log(1/\epsilon) + 6$.

Now we consider a simple example to illustrate the basic idea. Let $n = 2$, $p(00) = 1/2, p(01) = 1/4$, $p(10) = p(11) = 1/8$. Obviously, $h_{min}(p) = 1$ and $\Delta = (2 - 1)$. The map ϕ is constructed in two steps: first, "compress" the letters till $-\log p(a)$, that is, in our example, $00 \to 0$, $01 \to 10$ and $10 \to 110, 11 \to 111$. Secondly, randomise as follows: 00 uniformly $\to \{000, 001, 010, 011\}$, $10 \to \{100, 101\}$ and two last letters as $\{110\}$ and $\{111\}$ correspondingly. As a result, we obtain a set $\{0, 1\}^3$ subject to a uniform distribution whose min-entropy is equal to three, and hence $\Delta = 3 - 3 = 0$. Thus, the key length becomes 1 bit shorter, but the message length is longer. It is proved that such a "bloated" cipher is ϵ-entropically secure [11].

Obviously, the key length depends on the efficiency of the compression method, or code. Thus, in the case of known statistics (i.e., known p), the key length is $\Delta + 2log(1/\epsilon) + 2$, where Δ is 1 or 2 and depends on the compression code chosen. If p is unknown, but the messages are known to be generated by a Markov chain with known memory, then $\Delta = O(\log n)$ (and the key length is $O(\log n) + 2log(1/\epsilon)$ [11]).

2.4 Universal coding

The problem of constructing a single code for multiple probability distributions (information sources) is well known in information theory, and there are currently dozens of effective universal codes based on different ideas and approaches. It is worth noting that, at present, there are dozens universal codes, which are the basis for so-called archivers (e.g., ZIP). The first universal code for Bernoulli and Markov processes was proposed by Fitinghof [19], and then Krichevsky found an asymptotically optimal code for these processes [12,20]. Other universal codes include the PPM universal code [21], which is used together with the arithmetic code [22], the Lempel-Ziv (LZ) codes [23], the Burrows-Wheeler transformation [24], which is used together with the book-stack code (or MTF) [25] (see also also [26,27]), grammar codes [28,29] and some others [30–33].

The universal code c has to"compress" sequences $x = x_1...x_n$ that obey the distribution $p \in \mathbf{P}$ down to Shannon entropy p, that is $h_{Sh}(p)$, and the

difference between $E_p(|c(x)|) - h_{Sh}(p)$ is called redundancy $r(p)$ [12] (here E_p is the expectation and $|u|$ is the legth u). In [34], an algorithm was proposed to construct a code c_{opt} whose redundancy is minimal on $\mathbf{P}$, that is, $r_{p_{opt}} = \inf_{p\in\mathbf{P}} r(p)$. In [34] it was shown that $r_{p_{opt}}$ is equal to the capacity of a channel whose input alphabet is $\mathbf{P}$, whose output alphabet is the alphabet on which distributions from $\mathbf{P}$ are defined (in our case it is the alphabet $\{0,1\}^n$), and the lines of the channel matrix are probability distributions from $\mathbf{P}$ (see also [35] for the history of this discovery). This fact is important, because it allows us to use known methods to compute the channel capacity to find the optimal code.

In this paper, we will use the so-called Shtarkov maximum likelihood code c_{Sht} [36], whose construction is much simpler, and its redundancy is often close to that of the optimal code. This code is described as follows: first define

$$p_{max}(u) = \sup_{p\in\mathbf{P}} p(u), u \in \{0,1\}^n,\ S_{\mathbf{P}} = \sum_{u\in\{0,1\}^n} p_{max}(u),\ q(u) = p_{max}(u)/S_{\mathbf{P}}. \tag{5}$$

Clearly,

$$\forall u\ :\ p(u)/q(u) \le S_{\mathbf{P}}. \tag{6}$$

Shtakov proposed to build code c_{Sht} for which $|c_{Sht}(u)| = \lceil -\log q(u)\rceil$. (Such a code exists, see [37].)

Note that for a finite set P
$S_P \le |P|$ (In particular, this is true when P contains probability distributions corresponding to several languages).

3 The cipher

Now we are going to construct an ϵ-entropically secure cipher $c_{c\&r}$ for the case of unknown statistics, i.e., there exists some set of probability distributions $\mathbf{P}$ generating words from $\{0,1\}^n, n \ge 1$, and the constructed cipher should be applicable to messages obeying any $p \in \mathbf{P}$ with leakage no larger than ϵ. In short, we apply the general method from [11] to the probability distribution q (5). In detail, Alice wants to send messages $m \in \{0,1\}^n$ to Bob, and they both know in advance that m can obey any probability distribution p of the set of distributions $\mathbf{P}$. The cipher algorithm is as follows.

Constructing the cipher. We describe all calculations in the following steps:

i) compute the distribution q according to (5) and order the set $q(u), u \in \{0,1\}^n$. (Denote the ordered probabilities as $q_1, q_2, ..., q_N$, $N = 2^n$ and let $\nu(u) = i$ for which $q(u) = q_i$.)

ii) encode the "letters" $1, 2, ..., N$ with the distribution q by the trimmed Shannon code from [11] . Denote this code λ and note that

$$\forall i \, : \, |\lambda(i)| < -\log q_i + 2 \tag{7}$$

and λ is prefix-free, that is, for any i and j, $i \neq j$, neither $\lambda(i)$ is a prefix $\lambda(j)$, no $\lambda(j)$ is a prefix $\lambda(i)$ [11].

iii) build the following randomised map ϕ First, find $n^* = \max_i \lambda(i)$ and then define for $u \in \{0,1\}^n$,

$$\phi(u) = \lambda(\nu(u)) r_{|\lambda(\nu(u)|+1} ... \, r_{n^*} \, , \tag{8}$$

where r_j are equiprobable independent binary digits.

iv) For the desired leakage ϵ build $cipher_{rw-ds}$ with secret key length

$$\lceil \log S_{\mathbf{P}} \rceil + 2 \log(1/\epsilon) + \delta \, , \tag{9}$$

where $\delta = 2$ for ϵ-entropically secure cipher and $\delta = 6$ for ϵ- indistinguishable one.

It is worth noting that Alice and Bob (and Eve) can do all the calculations described independently of each other.

Use of the cipher. Suppose Alice and Bob have a randomly chosen secret key K, $|K| = k$, and Alice wants to send Bob a message m. To do this, she computes $cipher_{c\&r}(m, K)$, as described above, and sends it to Bob.

Bob receives the word $cipher_{c\&r}(m, K)$ and decrypts it with the key K. As a result he gets the word $\phi(m) = \lambda(\nu(m) r_{|\lambda(\nu(m)|+1} ... \, r_{n^*}$ whose prefix $\lambda(\nu(m))$ defines the message m (this is possible because λ is prefix-free).

The properties of this cipher are described in the following theorem.

Theorem 3. Suppose there is a family P of probability distributions defined on $\{0,1\}^n$ and some $\epsilon > 0$. If the described $cipher_{c\&r}$ is applied then

i) the $cipher_{c\&r}$ is ϵ-entropically secure with secret key length $\lceil \log S_{\mathbf{P}} \rceil + 2 \log(1/\epsilon) + 2$ and

ii) the $cipher_{c\&r}$ is ϵ–indistinguishable with secret key length $\lceil \log S_{\mathbf{P}} \rceil + 2 \log(1/\epsilon) + 6$.

Proof. For any $p \in \mathbf{P}$ the random map ϕ defines a probability distribution $\pi_p(v), v \in \{0,1\}^*$ as follows: for any $u \in \{0,1\}^n$ and $v \in \phi(u)$

$$\pi_p(v) = p(u)2^{-(n^* - |\lambda(\nu(u)|)}\,,$$

see (8). From definitions ϕ and (8), (7) we obtain

$$\pi_p(v) = p(m)2^{-(n^* - |\lambda(\nu(m)|)} \le p(m)2^{-(n^* - (\log q_{\nu(m)} + 2))}$$

for any $m \in \{0,1\}^n$ and $v \in \phi(m) \subset \{0,1\}^{n^*}$. Then

$$-\log \pi_p(v) \ge -\log p(m) - (n^* - (\log q_{\nu(m)} + 2)) \ge$$

$$\log S_{\mathbf{P}} - \log q_{\nu(m)} - (n^* - (\log q_{\nu(m)} + 2)) = \log S_{\mathbf{P}} + 2 - n^*$$

for any m and $v \in \phi(m) \subset \{0,1\}^{n^*}$. So, $h_{min}(\pi_p) = \min_{v \in \{0,1\}^{n^*}} -\log \pi_p(v) \ge \log S_{\mathbf{P}} + 2 - n^*$ and, hence, $\sup_{p \in \mathbf{P}}(n^* - h_{min}(\pi_p)) \le \log S_{\mathbf{P}} + 2$. From (4) (Theorem 2) and the description of the cipher (9) we can see that the $cipher_{c\&r}$ is

i) ϵ-entropically secure with a secret key of length $\lceil \log S_{\mathbf{P}} \rceil + 2\log(1/\epsilon) + 4$ and

ii) ϵ-indistinguishable with a secret key of length $\lceil \log S_{\mathbf{P}} \rceil + 2\log(1/\epsilon) + 8$.

4 Conclusion

We described the cipher for a family of probability distributions $\mathbf{P}$ defined on the set $\{0,1\}^n, n \ge 1$, for which the length of the secret key does not depend directly on n, but depends on $\mathbf{P}$. For example, if $\mathbf{P}$ is finite, the key length is less than $\log |\mathbf{P}| + 2\log(1/\epsilon) + O(1)$ and hence independent of n. This example includes the case where one needs to have the same cipher for texts written in different languages. Here, the size of the set $\mathbf{P}$ is equal to the number of languages. Thus, in some practically interesting cases, the extra length of the secret key is quite small.

References

[1] Shannon C. E. Communication theory of secrecy systems. The Bell system technical journal. 1949 Oct; 28(4):656-715.

[2] Juels A, Ristenpart T. Honey encryption: Security beyond the brute-force bound. In Annual international conference on the theory and applications of cryptographic techniques 2014 May 11 (pp. 293-310). Springer, Berlin, Heidelberg.

[3] Russell A, Wang H. How to fool an unbounded adversary with a short key. IEEE Transactions on Information Theory. 2006 Mar 6;52(3):1130-40.

[4] J Jaeger, T Ristenpart, Q Tang. Honey encryption beyond message recovery security. IACR Cryptology ePrint Archive; 2016.

[5] Dodis Y., Smith A. Entropic security and the encryption of high entropy messages. In: Theory of Cryptography Conference 2005 Feb 10 (pp. 556-577). Springer, Berlin, Heidelberg.

[6] F. du Pin Calmon, M. Medard, L. M. Zeger, J. Barros, M. M. Christiansen, and K. R. Duffy. Lists that are smaller than their parts: A coding approach to tunable secrecy. In 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012, October 1-5, 2012, pp. 1387?1394. IEEE, 2012.

[7] Calmon F. D. Information-theoretic metrics for security and privacy (Doctoral dissertation, Massachusetts Institute of Technology), 2015.

[8] Ryabko, B. A simply realizable ideal cryptographic system. Problems of Information Transmission, 36, (2000), no. 1, pp. 84-89. (see also IACR Cryptology ePrint archive, report 2001/046).

[9] Ryabko, B. The Vernam Cipher Is Robust to Small Deviations from Randomness. Problems of Information Transmission, 2015, 51(1), pp. 82-86.

[10] Li X., Tang Q., Zhang Z. Fooling an Unbounded Adversary with a Short Key, Repeatedly: The Honey Encryption Perspective. In 2nd Conference on Information-Theoretic Cryptography (ITC 2021) 2021. Schloss Dagstuhl-Leibniz-Zentrum Informatik.

[11] Ryabko, B. Unconditionally secure short key ciphers based on data compression and randomization. Des. Codes Cryptogr., pp.1-12, 2023.

[12] Krichevsky R. *Universal Compression and Retrival*. Kluver Academic Publishers, 1993.

[13] Shkel YY, Poor HV. A compression perspective on secrecy measures. IEEE Journal on Selected Areas in Information Theory. 2021 Feb 2;2(1):163-76.

[14] Bloch M, Günlü O, Yener A, Oggier F, Poor HV, Sankar L, Schaefer RF. An overview of information-theoretic security and privacy: Metrics, limits and applications. IEEE Journal on Selected Areas in Information Theory. 2021 Mar 17;2(1):5-22.

[15] Ryabko B., Fionov A. Cryptography in the Information Society. - World Scientific Publishing. - 2020. - 280 p.

[16] Gunther C. G. A universal algorithm for homophonic coding. InWorkshop on the Theory and Application of Cryptographic Techniques 1988 May 25 (pp. 405-414). Springer, Berlin, Heidelberg.

[17] Naor J, Naor M. Small-bias probability spaces: Efficient constructions and applications. InProceedings of the twenty-second annual ACM symposium on Theory of computing 1990 Apr 1 (pp. 213-223).

[18] Alon N, Goldreich O, Håstad J, Peralta R. Simple constructions of almost k-wise independent random variables. Random Structures & Algorithms. 1992;3(3):289-304.

[19] Fitingof B. M. Optimal coding in the case of unknown and changing message statistics, Problemy Peredachi Informatsii, 2(2), 3-11, 1966

[20] Krichevsky R. A relation between the plausibility of information about a source and encoding redundancy. Problems Inform. Transmission. 1968;4(3):48-57.

[21] J. Cleary and I. Witten, "Data compression using adaptive coding and partial string matching," *IEEE transactions on Communications*, vol. 32, no. 4, pp. 396–402, 1984.

[22] J. Rissanen and G. G. Langdon, "Arithmetic coding," *IBM Journal of research and development*, vol. 23, no. 2, pp. 149–162, 1979.

[23] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on information theory*, vol. 23, no. 3, pp. 337–343, 1977.

[24] M. Burrows and D. J. Wheeler, "A block-sorting lossless data compression algorithm," 1994.

[25] B. Y. Ryabko, "Data compression by means of a "book stack"," *Problemy Peredachi Informatsii*, vol. 16, no. 4, pp. 16–21, 1980.

[26] J. Bentley, D. Sleator, R. Tarjan, and V. Wei, " A locally adaptive data compression scheme," *Communications of the ACM*, vol. 29, no. 4, pp. 320–330, 1986.

[27] B. Ryabko, N. R. Horspool, G. V. Cormack, S. Sekar, and S. B. Ahuja, "Technical correspondence," *Communications of the ACM*, vol. 30, no. 9, pp. 792–797, 1987.

[28] J. C. Kieffer and E.-H. Yang, "Grammar-based codes: a new class of universal lossless source codes," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 737–754, 2000.

[29] E.-H. Yang and J. C. Kieffer, "Efficient universal lossless data compression algorithms based on a greedy sequential grammar transform. i. without context models," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 755–777, 2000.

[30] M. Drmota, Yu. Reznik, and W. Szpankowski, " Tunstall code, Khodak variations, and random walks," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2928–2937, 2010.

[31] G. Louchard, W.Szpankowski, "Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm", IEEE Transactions on Information Theory. vol. 41, no. 2, pp. 478-488, 1995.

[32] B. Ryabko, "Twice-universal coding," *Problems of Information Transmission*, vol. 3, pp. 173–177, 1984.

[33] Y. A. Reznik Coding of Sets of Words. 2011 Data Compression Conference, IEEE, 2011.

[34] Ryabko, B. Coding of a source with unknown but ordered probabilities. Problems Inform. Transmission 15 (1979), no. 2, 134–138;

[35] Ryabko, B. Comments on: "A source matching approach to finding minimax codes", IEEE Trans. Inform. Theory 27 (1981), no. 6, 780–781.

[36] Shtar'kov YM. Universal sequential coding of single messages. Problemy Peredachi Informatsii. 1987;23(3):3-17.

[37] T. M. Cover and J. A. Thomas, *Elements of information theory.* New York, NY, USA: Wiley-Interscience, 2006.