Designs, Codes and Cryptography (2023) 91:2201–2212
https://doi.org/10.1007/s10623-023-01195-8

Unconditionally secure short key ciphers based on data compression and randomization

Boris Ryabko[1]

Received: 30 July 2022 / Revised: 2 February 2023 / Accepted: 3 February 2023 /
Published online: 27 February 2023

Abstract
We consider the problem of constructing an unconditionally secure cipher for the case when the key length is less than the length of the encrypted message. (Unconditional security means that a computationally unbounded adversary cannot obtain information about the encrypted message without the key). In this article, we propose a cipher based on data compression and randomisation in combination with entropically-secure encryption and apply it to the following two cases: (i) the statistics of encrypted messages are known; and (ii) statistics are unknown, but messages are generated by a Markov chain with known memory (or connectivity). In both cases, the length of the secret key is negligible compared to the length of the message.

Keywords Information theory · Cryptography · Perfect security · Entropic security · Entropically-secure symmetric encryption scheme · Indistinguishability · Data compression · Randomisation · Shannon code · Fitingof code

Mathematics Subject Classification 94A60

1 Introduction

The concept of unconditional secrecy was presented in the seminal article by Shannon, where he also showed that a one-time pad (or Vernam cipher) is unconditionally secure [23]. In particular, unconditional secrecy means that a computationally unbounded adversary cannot obtain any information about an encrypted message without a key. It is clear that this property is highly desirable, but if the one-time pad is used to encrypt a message, the length of the key used must be at least the length of the message (or, more precisely, its Shannon entropy). This requirement is too strict for many applications, and there are many

Communicated by A. Winterhof.

✉ Boris Ryabko
boris@ryabko.net

[1] Federal Research Center for Information and Computational Technologies, Novosibirsk State University, Novosibirsk, Russia

approaches for creating secure ciphers with short or low entropy keys, see [3, 5, 6, 11, 12, 15, 17, 18]. One such approach was suggested by Shannon in [23], who described ideal cipher systems where a computationally unbounded adversary "does not obtain a unique solution to the cipher but is left with many alternatives, all of reasonable probability". He built a theory of ideal ciphers and described some of them for the case when the probability distribution of encrypted messages is known. Later, ideal systems were proposed for the case of unknown statistics, see [17].

An interesting new way of building secure systems with short keys is the so-called entropic security, proposed by Russell and Wang [15] and developed by Dodis and Smith [5]. This cipher uses the entropy of the original message in such a way that the key length should be roughly the difference between the message length and its min-entropy (the exact definition will be given below). (Note that the use of the entropy message to improve the strength of the cipher was also used in [17, 18].)

The notion of an entropically-secure symmetric encryption scheme is extremely important for cryptography because one can construct this scheme with a key shorter than the length of the input. In a sense, this circumventins Shannon's lower bound on the key length.

Another way to construct a short key cipher is the so-called honey cipher, proposed by Juels and Ristenpart [12] and developed by Jaeger et al. [11]. In some ways, the honey cipher is like the ideal cipher: a computationally unlimited adversary has many possible highly probable decryptions. Li et al. [14] developed and combined the ideas of the honey cipher and the entropically-secure ciphers to create a new class of easily implementable short key codes. In a sense, the idea of preprocessing an original message in order to increase its entropy is being developed and widely used in their methods.

Data compression and randomization are two methods of preprocessing the original message that have been used for centuries in cryptography [10, 23]. Moreover, homophonic coding can be used to compress and randomize together [10, 20]. The goal of both transformations is to make the probability distribution of the original messages closer to the uniform one (see an overview in [1]). Interestingly, both transformations have been successfully applied to some cryptographic problems: they were used to extract randomness [7, 21, 24] and to build an ideal steganographic system [22].

In this article, we combine entropically-secure encryption with the suggested compression and data randomization techniques and apply it to the following two cases: (i) the statistics of encrypted messages are known; and (ii) statistics are unknown, but messages are generated by a Markov chain with known memory (or connectivity). In the first case the key length is a constant, whereas in the second case it is $O(\log n)$, where n is the length of the message. (But in both cases, the length of the secret key depends on the required security level). This makes it possible to apply an entropically-secure cipher so that the key length is independent of the entropy or the length of the message.

2 Definitions and preliminaries

We consider the problem of symmetric encryption, when there are two parties Alice and Bob and Alice wants to securely transmit a message M to Bob, where $M \in \{0, 1\}^n$, $n \geq 1$, obeys a certain probability distribution p defined on the set $\{0, 1\}^n$. Alice and Bob have a shared secret key $K = K_1 \dots K_k$, which can be much shorter than the length of M, that is $k \ll n$. Alice encrypts M with K and possibly some random bits, and obtains the word $cipher(M, K)$. Then she sends it to Bob, who decrypts the received message and obtains

M. In addition, there is a computationally unlimited adversary Eve who does not know M and K, but knows the probability distribution p and wants to find some information about M based on the encrypted message.

Let Λ be a finite alphabet and $\mathbf{P}$ be a set of probability distributions defined on $A = \Lambda^n$ where $n \geq 1$. By definition, the min-entropy of $p \in \mathbf{P}$ is as follows

$$h_{min}(p) = -\log \max_{a \in A} p(a)\,. \tag{1}$$

(Here and below $\log = \log_2$.)

In this article we will consider ciphers that can be applied to messages obeying not just one particular distribution, but any distribution from some family. We will therefore define the following two families of probability distributions: $\mathbf{P}_{min}(h) = \{p : h_{min}(p) \geq h\}$, and $\mathbf{P}_{markov}(m)$ which contains stationary ergodic Markov chains with memory, or connectivity, $m, m \geq 0$. (The definition can be found in [2, 19] and Appendix).

Russell and Wang [15] suggested a definition of the entropic security which was generalised by Dodis and Smith [5] and can be formulated as follows: A probabilistic map Y is said to hide all functions on Λ^n with leakage ϵ if, for every adversary B, there exists some adversary $\hat{B}$ (who does not know $Y(M)$) such that for any distribution from $\mathbf{P}$, all functions f, and any $M \in \Lambda^n$

$$|\, Pr\{B(Y(M)) = f(M)\} - Pr\{\hat{B}() = f(M)\}\,| \leq \epsilon. \tag{2}$$

(note that $\hat{B}$ does notknow $Y(M)$ 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 $f : \Lambda^n \to \{0, 1\}^*$.

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

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

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

(So, B and $\hat{B}$ obtained the same result, but B estimates the probability based on $cipher(M, K)$, whereas $\hat{B}$ does it without knowledge of $cipher(M, K)$). So, the entropic security (2) can be considered as a generalisation of the Shannon's perfect secrecy.

The following theorem of Dodis and Smith [5] is a generalisation of the results of Russell and Wang [15].

Theorem 1 [5] *Let there be an alphabet* $\{0, 1\}^n, n > 0$, *with a probability distribution* p *whose min-entropy is not less then* h *(that is,* $p \in \mathbf{P_{min}}$ (h) *with* $\Lambda = \{0, 1\}$*). Then there exists an efficient* 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 $cipher_{rw-ds}$ *(A description of one of these ciphers from [5] is given in the Appendix). It is important to note that one cipher is applied for any distribution whose min-entropy is not greater than* h.

Another important notion is that of indistinguishability:

Definition 2 [5] A randomised map $Y : \{0,1\}^n \to \{0,1\}^n$, $n \geq 1$, is ϵ-indistinguishable for some family of destributions **P** 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

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

Dodis and Smith [5] showed that entropic security and indistinguishability are deeply connected for distributions with bounded min-entropy:

Theorem 2 [5] *Let Y be a randomised map with inputs of length n bits. Then*

1. *ϵ-entropic security for $\mathbf{P_{min}}$ (t) implies 4ϵ-indistinguishability for $\mathbf{P_{min}}$ $(t-1)$ and*
2. *ϵ-indistinguishability for $\mathbf{P_{min}}$ $(t-2)$ implies $\epsilon/8$ -entropic security for $\mathbf{P_{min}}$ $(t-2)$, when $2\log(1/\epsilon) + 1 \geq t$.*

It is worth noting that the same cipher (or random map Y) is ϵ-entropically secure and ϵ-indistinguishable, and both concepts are equivalent up to small parameter changes. Moreover, one cipher (and one Y) fits all distributions from $\mathbf{P_{min}(t)}$.

3 Suggested cipher: general construction

Suppose there is a set **M** of n-letter messages from Λ and a family of probability distributions **P** on **M**. We can see from Theorem 1 that the choice of the length k of the key K depends significantly on the min-entropy of the probability distribution; specifically, $k = n - h_{min} + 2\log(1/\epsilon) + 2$.

So, the following observation seems to be natural: transform the set **M** into a larger set of longer messages $\mathbf{M}^*$ in such a way that the minimum entropy of the transformed distribution becomes closer to the message length in $\mathbf{M}^*$. Then, apply the entropically-secure cipher $cipher_{rw-ds}$ to messages from $\mathbf{M}^*$. It turns out that the resulting cipher for the original set **M** is also entropy-secure with a short key.

The proposed cipher is based on this observation and is described as follows. Suppose that there is a set **M** of n-letter messages with a probability distribution $p \in \mathbf{P}$.

Preliminary stage Alice and Bob build such a random map ϕ of $\Lambda^n \to \{0,1\}^{n^*}$, $n^* \geq n$, that

i. there exists such a map $\phi^{-1} : \{0,1\}^{n^*} \to \Lambda^n$ that $\phi^{-1}(\phi(m)) = m$ for any $m \in \mathbf{M}$, and
ii. for any $p \in \mathbf{P}$ the min-entropy of the corresponding probability distribution π_p on $\mathbf{M}^*$ is $n^* - \Delta$, where Δ is significantly less than n^*. (That is, ϕ converts p to π_p. Formally, for $v \in \mathbf{M}^*$, $\pi_p(v) = Pr\{\phi(u) = v\}$, $u \in \mathbf{M}$, and u obeys the distribution p).

Message encryption Suppose that Alice wants to cipher a message $m \in \mathbf{M}$ with a leakage ϵ. She calculates $z = cipher_{rw-ds}$ $(\phi(m))$ with key length $\Delta + 2\log(1/\epsilon) + 2$ and sends z to Bob. To decrypt, Bob computes ϕ^{-1} $(decipher_{rw-ds}(z))$.

The random map ϕ uses data compression and randomisation, so that we denote this cipher as $cipher_{c\&r}$.

Theorem 3 *Let there be a set of messages* $\mathbf{M} \subset \Lambda^n$, $n > 0$, *and a probability distribution* p *on* $\mathbf{M}$, $p \in \mathbf{P}$. *Let* $\epsilon > 0$ *and assume that the described cipher* $cipher_{c\&r}$ *applies to messages from* $\mathbf{M}$ *with a secret key of length* $\Delta + 2\log(1/\epsilon) + 2$ *and properties (i) and (ii) are valid.*

Then this cipher is ϵ*-entropically secure, that is, for any function* $A : \{0,1\}^* \to \{0,1\}^*$, *any* $f : \Lambda^n \to \{0,1\}^*$ *and* $M \in \mathbf{M}$

$$|Pr\{A(cipher_{c\&r}(M) = f(M)\} - Pr\{\hat{A}() = f(M)\}| \leq \epsilon,$$

where $\hat{A}$ *does not use the value* $cipher_{c\&r}(M)$.

Proof Taking into account that the min-entropy of π_p, $p \in \mathbf{P}$, equals $n^* - \Delta$, we see from Theorem 1 that for any function g

$$|Pr\{A(cipher_{rw-ds}(v) = g(v)\} - Pr\{\hat{A}() = g(v)\}| \leq \epsilon,$$

where v is a random variable with distribution π_p on $\{0,1\}^{n^*}$, g is any function defined on $\{0,1\}^{n^*}$ ($g : \{0,1\}^{n^*} \to \{0,1\}^*$) and $\hat{A}()$ does not depend on $cipher_{rw-ds}(v)$. For any function $f : \Lambda^n \to \{0,1\}^*$, any $u \in \mathbf{M}$ and any $v = \phi(u)$ define the function $g(v) = f(\phi^{-1}(v))(= f(u))$ (so, $g(v)$ is defined if $\pi_p(v) > 0$). The last inequality is valid for this function g, hence

$$|Pr\{A(cipher_{rw-ds}(\phi(u)) = f(u)\} - Pr\{\hat{A}() = f(u)\}| \leq \epsilon.$$

Taking into account that $cipher_{c\&r}(u) = cipher_{rw-ds}(\phi(u))$ and $f(\phi^{-1}(v)) = f(u)$, we can see from the last inequality that

$$|Pr\{A(cipher_{c\&r}(u)) = f(u)\} - Pr\{\hat{A}() = f(u)\}| \leq \epsilon\,.$$

The theorem is proven. □

We can see from Theorem 2 that the cipher $cipher_{c\&r}$ is 2ϵ indistinguishable. Below we prove directly that $cipher_{c\&r}$ is ϵ indistinguishable with $k = \Delta + 2log(1/\epsilon) + 6$.

From Theorem 2 we know that, in fact, the indistinguishability is equal to the entropic security, and it holds for $cipher_{rw-ds}$, but we are interested in the indistinguishability of the $cipher_{c\&r}$. The following theorem establishes this.

Theorem 4 *Let there be set of messages* $\mathbf{M} \subset \Lambda^n$, $n > 0$, *and a probability distribution* p *on* $\mathbf{M}$, $p \in \mathbf{P}$. *Let* $\epsilon > 0$ *and assume that the described cipher* $cipher_{c\&r}$ *applies to messages from* $\mathbf{M}$ *with a secret key of length* $\Delta + 2\log(1/\epsilon) + 6$ *and properties (i) and (ii) are valid.*

Then this cipher is ϵ-indistinguishable.

Proof In order to prove it suppose that $cipher_{wr-ds}$ is applied to the words from the set $\phi(\mathbf{M}) \subset \{0,1\}^{n^*}$ in such a way that it is $(1, \epsilon/4)$ entropically secure, where the length of the key equals $\Delta + 2\log(1/(\epsilon/4)) + 2 = 2\log(1/\epsilon) + 6$. From Theorem 2 we can see that this cipher is ϵ indistinguishable, that is, $SD(cipher_{rw-ds}, G) \leq \epsilon$, where G is a random variable on $\{0,1\}^{n^*}$ which is independent on $cipher_{rw-ds}$.

Define $U_a = \{cipher_{rw-ds}(\phi(a))\}\}$ and let $G'(v)$ be defined as follows:

$$Pr\{G' = v\} = \sum_{w \in U_v} Pr\{G = w\}.$$

The following chain of equalities and inequalities is based on these definitions and the triangle inequality for L_1:

$$
\begin{aligned}
&SD(cipher_{c\&r}, G') = \\
&\quad \frac{1}{2} \sum_{u \in \Lambda^n} |Pr\{cipher_{c\&r} = u\} - Pr\{G' = u\}| = \\
&\quad \frac{1}{2} \sum_{v \in \Lambda^n} | \sum_{w \in U_v} Pr\{cipher_{rw-ds} = w\} - Pr\{G = w\}| \le \\
&\quad \frac{1}{2} \sum_{v \in \Lambda^n} \sum_{w \in U_v} |Pr\{cipher_{rw-ds} = w\} - Pr\{G = w\}| = \\
&\quad \frac{1}{2} \sum_{w \in \{0,1\}^{n^*}} |Pr\{cipher_{rw-ds} = w\} - Pr\{G = w\}| = \\
&\quad SD(cipher_{rw-ds}, G) \le \epsilon \, .
\end{aligned}
$$

So, $SD(cipher_{c\&r}, G') \le \epsilon$.

Theorem is proven. □

4 Applying $cipher_{c\&r}$ to messages generated by a source with known statistics

In this part we apply $cipher_{c\&r}$ to messages from a set $\{0, 1\}^n$ which obey a distribution p, whereas p is known to Alice, Bob and Eve. Let $A \subset \{0, 1\}^n$ be a set of messages with non-zero probabilities. In this part, it will be convenient to call $A = \{a_1, \ldots, a_L\}$ alphabet and a_i letters, since we will consider letter-wise codes. Clearly,

$$\log L \le n \, . \tag{4}$$

We first describe the preliminary stage in which we use the Shannon code [4] to compress the messages and then randomisation.

4.1 Lossless codes

4.1.1 Shannon code and its generalisations

Suppose, $a_1, \ldots, a_L \in A$ are ordered in such a way that $p(a_1) \ge p(a_2) \ge \cdots \ge p(a_L) > 0$. Define $Q_1 = 0$, $Q_t = \sum_{i=1}^{t-1} p(a_i), t = 2, \ldots, L$, and let $\hat{Q}_i, t = 1, \ldots, L$, be a presentation of Q_i in binary system as an infinite $\{0, 1\}$ word and without the initial 0. (if there are two such presentations then the presentation with finite number of ones is used). That is, $1/2 = 100000\ldots$, $1/3 = 010101\ldots$. The codeword $\hat{\lambda}(a_i)$ for symbol a_i is chosen to be the first $\lceil \log(1/p(a_i) \rceil$ binary digits in $\hat{Q}_i, i = 1, \ldots, L$. It is clear that,

$$|\hat{\lambda}(a_i)| = \lceil \log(1/p(a_i)) \rceil \, . \tag{5}$$

For example, let $A = \{a_1, a_2, a_3\}$ and $p(a_1) = 13/16$, $p(a_2) = 1/8$, $p(a_3) = 1/16$. Then, $\hat{\lambda}(a_1) = 0$, $\hat{\lambda}(a_2) = 110$, $\hat{\lambda}(a_3) = 1111$. Clearly, these codewords can be made shorter as follows: $\lambda(a_1) = 0$, $\lambda(a_2) = 10$, $\lambda(a_3) = 11$. This procedure for removing extra digits can be described using binary trees. It is known that the Shannon code can be represented as a binary tree, the branches of which correspond to codewords. In this tree, the left child is marked with 0, and the right child is 1. If some node has one child, it is removed, and the corresponding digit from the corresponding codeword is also removed. The obtained code we denote as λ_{Sh} and derive from (5) the following:

$$|\lambda_{Sh}(a_i)| \leq \lceil \log(1/p(a_i)) \rceil \leq \log(1/p(a_i)) + 1. \tag{6}$$

Also, it is known that the set of codewords $\lambda_{Sh}(a_1), \ldots, \lambda_{Sh}(a_L)$ is prefix-free. (Recall that, by definition, a set of words U is prefix-free if for any $u, v \in U$ neither u is a prefix of v nor v is a prefix of u.)

Note that, for any a prefix-free code λ' and any sequence $x_1 x_2 \ldots x_n$ from A, $n \geq 1$, the encoded sequence $\lambda'(x_1)\lambda'(x_2)\ldots\lambda'(x_n)$ can be decoded to $x_1 x_2 \ldots x_n$ without errors. Such a code λ' is called lossless code. Hence, any prefix-free code is a lossless one.

Note the the "initial" code $\hat{\lambda}(a_i)$ has the same properties as a modified λ_{Sh}, that is, it is the prefix-free and (6) is valid. (That is why we do not describe the transformation of $\hat{\lambda}$ to λ_{Sh} in detail and do not estimate its complexity.)

4.1.2 Trimmed codes

Let λ be a lossless code for elements from A. Consider the following probability distribution $p(a_1) = 1/2$, $p(a_2) = 1/4, \ldots, p(a_{L-1}) = p(a_L) = 2^{-(L-1)}$. From the description of the Shannon code we can see that $|\lambda_{Sh}(a_L))| = L - 1$.

In applications, the complexity of the cipher will largely depend on the lengths of the codewords. Thus, it will be convenient to use codes for which the length of the code of any letter does not exceed $\lceil \log L \rceil + 1$ for any probability distribution (instead of $L - 1$ as in the previous example). It is also worth noting that it will be shown later that one extra bit of the length of the codeword can add at most 1 extra bit of the length of the encryption key. We call such codes as trimmed and define one of them as follows: if λ is a code then

$$\lambda^{tr}(a_i) = \begin{cases} 0\,\lambda(a_i) & \text{if } |\lambda(a_i)| \leq \lceil \log L \rceil \\ 1\,bin_{\lceil \log L \rceil}(i) & \text{if } |\lambda(a_i)| > \lceil \log L \rceil, \end{cases} \tag{7}$$

where $bin_{\lceil \log L \rceil}(i)$ is a binary presentation of i whose length is $\lceil \log L \rceil$. (For example, $bin_3(3) = 011$). We see that the maximal codeword length is not greater than $\lceil \log L \rceil + 1$. Also, note that for any prefix-free code the maximal codeword length is not less than $\lceil \log L \rceil$.

Let us explain how to decode. First, the decoder reads the first binary letter. If it is 0, the decoder uses the codeword of the code λ in order to find the encoded letter. If the first letter is 1, the next $\lceil \log L \rceil$ letters contain the binary decomposition of i, i.e. the letter is a_i.

If the trimmed code is built based on the Shannon code, from (7) and (6) we obtain

$$|\lambda^{tr}_{Sh}(a_i)| \leq \lceil \log(1/p(a_i)) \rceil + 1 < \log(1/p(a_i)) + 2. \tag{8}$$

4.2 Randomised prefix-free codes

Let λ be a prefix-free code for the alphabet A and

$$n^* = \max_{i=1,\ldots,L} |\lambda(a_i)|.$$

The randomised code ϕ_λ maps letters from the alphabet A to the set $\{0, 1\}^{n^*}$ defined as follows.

$$\phi_\lambda(a_i) = \lambda(a_i)\, r^i_{|\lambda(a_i)|+1} r^i_{|\lambda(a_i)|+2} \cdots r^i_{n^*}, \tag{9}$$

where $r^i_{|\lambda(a_i)|+1}, r^i_{|\lambda(a_i)|+2}, \ldots, r^i_{n^*}$ uniformly distributed and independent random bits (for all i). Let us define a probability distribution π_λ on $\{0,1\}^{n^*}$ as follows:

$$\pi_\lambda(y_1 y_2 \ldots y_{n^*}) = p(a_i) 2^{-(n^* - |\lambda(a_i)|)} \\ \text{if} \quad y_1 y_2 \ldots y_{|\lambda(a_i)|} = \lambda(a_i). \tag{10}$$

If for some $y = y_1 \ldots y_{n^*}$ any $\lambda(a_i)$ is not a prefix of y, then $\pi_\lambda(y) = 0$.

Claim 1 $h_{min}(\pi_\lambda) = n^* - \max_{i=1,\ldots,L}(|\lambda(a_i)| - \log(1/p(a_i)))$. *In particular,*

$$h_{min}(\pi_{\lambda_{Sh}}) > n^* - 1, \;\; h_{min}(\pi_{\lambda^{tr}_{Sh}}) > n^* - 2. \tag{11}$$

Here the first equation follows from the definition of the min-entropy and (10), whereas (11) follows from (6) and (8).

4.3 The cipher for known statistics

Now we can apply the cipher from the part 4 (see message encryption) with $\phi = \phi^{tr}_{Sh}$ and $\pi = \pi_{\lambda^{tr}_{Sh}}$. From (11) we can see that $\Delta = 2$ and applying Theorems 3 and 4 and the estimate we obtain

Claim 2 *Let there be set of messages from* $\{0,1\}^n, n > 0$, *obey a known probability distribution* p *and* $\epsilon > 0$. *Let the cipher* $cipher_{c\&r}$ *be applied*

(i) *with a secret key of the length* $4 + 2\log(1/\epsilon)+$ *bits. Then this cipher is* ϵ*-entropically secure.*
(ii) *If the key length* k *equals* $k = 8 + 2log(1/\epsilon)$ *bits, the cipher is* ϵ*-indistinguishable.*

Comment In this section, we described the Shannon code, for which the letters of the alphabet must be arranged in descending order of probabilities. Sometimes it can be a time consuming operation. In such a case, one can use the Gilbert-Moore code [9], which can be used for unordered probabilities, but its code length is one bit longer than the Shannon code, i.e. the code length of a_i can be $\log(1/p(a_i)) + 2$. For this code, we can also use a trimmed version. In both cases, the use of the Gilbert–Moore code may add one extra bit to the key length.

5 Messages generated by Markov chains with unknown statistics

As in the previous part, we perform encryption in three steps: compress, randomise, and apply $cipher_{rw-ds}$. For compression, we apply the so-called universal codes, which are used for unknown statistics.

5.1 Universal coding

First, we consider the simplest case where the alphabet is $\{0,1\}^n, n \geq 1$, and letters are generated by some i.i.d. source μ and $\mu(0), \mu(1)$ are unknown. The goal is to build a lossless code which "compresses" n-letter sequences in such a way that the average length (per letter) of the compressed sequence is close to the Shannon entropy $h(\mu)$, which is the lower limit of the codeword length (lossless code is such that the encoded messages can be decoded without errors and $h(\mu) = -(\mu(0)\log\mu(0) + (1-\mu(0))\log(1-\mu(0))$).

The first universal code was invented by Fitingof [8] and we use this code as a part of the suggested entropically secure cipher. In order to describe this code we consider any word $v \in \{0,1\}^n$ and denote by ν the number of ones in v and let S_ν be the set of n-length words with ν ones. Fitingof proposed to encode the word v by two subwords u (prefix) and w (suffix), where u is the binary notation of an integer ν and w is the index of the word v in the subset S_ν. It is assumed that the words in S_ν are ordered 0 to $(|S_\nu|-1)$ (say, lexicographically) and the lengths of u and w are equal to $\lceil \log(n+1) \rceil$ and $\lceil \log |S_n| \rceil$, respectively. For example, for $n = 3$, $v = 100$ we obtain $\nu = 1$, $u = 01$, $w = 10$. Clearly, this code is prefix-free.

If we denote the Fitingof code by $code_F$ we obtain from its description

$$|code_F(v)| = \lceil \log(n+1) \rceil + \lceil \log |S_\nu| \rceil \,. \tag{12}$$

For this code the ability to compress messages is based on the simple observation that probabilities of all messages from S_ν are equal for any distribution μ and, hence, $\mu(v) \le 1/|S_\nu|$ for μ and any word $v \in S_\nu$. From this inequality and (12) we obtain

$$|code_F(v)| \le \log(n+1) + 2 + \log(1/\mu(v)) \,. \tag{13}$$

(Let's explain the name “universal code.” Clearly, the average code-length $E_\mu(|code_F|)$ is not greater than $\log(n+1)+2+nh(\mu)$ and, hence, the average length per letter $E_\mu(|code_F|)/n$ is not grater than $h(\mu) + (\log(n+1)+2)/n)$. We can see that $E_\mu(|code_F|)/n \to h(\mu)$ if $n \to \infty$. So, one code compresses sequences generated by any μ, that is, the code is universal).

The Fitingof code described generalizes to i.i.d. processes with any finite alphabet Λ, as well as to Markov chains with memory or connectivity m, based on the same method as for binary i.i.d. [13]. Namely, the set of all n-letter words is divided into subsets of equiprobable words, and the code of any word is represented by a prefix and a suffix, where the prefix contains the number of the set with equiprobable words which contains the encoded one, and the prefix is the number in this set. It can be shown that the number of sets with equiprobable words is bounded above by $(|\Lambda|-1)|\Lambda|^m$ [8, 13], and similarly (13) we can deduce that

$$|code_F(v)| \le \log((|\Lambda|-1)|\Lambda|^m) + 2 + \log(1/\mu(v)) \,. \tag{14}$$

It is important to note that there exists an algorithm to find the codewords which is based on method of fast calculation of numbers in S_ν, see [16]. The complexity of this algorithm is $O(n \log^3 n \log\log n)$.

5.2 Randomisation

As with known statistics, we randomise compressed messages to construct random maps of ϕ and ϕ^{-1} for which $\phi^{-1}(\phi(u)) = u$ for any message u (see "preliminary stage"). This method is similar from the part from 4.2.

Let $n^* = \max_{w \in \Lambda^n}\{|code_F(w)|\}$. The randomized code ϕ_F maps elements from Λ^n to the set $\{0,1\}^{n^*}$ defined as follows:

$$\phi_F(w) = code_F(w)\, r^i_{|code_F(w)|+1} r^i_{|code_F(w)|+2} ... r^i_{n^*} \,, \tag{15}$$

where $r^i_{|code_F(w)|+1}, r^i_{|code_F(w)|+2}, ..., r^i_{n^*}$ are uniformly distributed and independent random bits (for all i).

Let us define the probability distribution $\pi_{F,\mu}$ on $\{0,1\}^{n^*}$ as follows:

$$\pi_{F,\mu}(y_1 y_2 ... y_{n^*}) = \mu(v) 2^{-(n^* - |code_F(v)|)} \text{ if } y_1 y_2 ... y_{|code_F(v)|} = code_F(v). \tag{16}$$

If for some $y = y_1 \dots y_{n^*}$ any $v \in \Lambda^n$, $code_F(v)$ is not a prefix of y, then $\pi_{F,\mu}(y) = 0$.

Let us estimate the min-entropy of the distribution $\pi_{F,\mu}$. From this equation and the definition of the min-entropy we obtain the following:

$$h_{min}(\pi_{F,\mu}) = n^* - \max_{u \in \Lambda^n}(|code_F(u)| - \log(1/\mu(u))\,. \tag{17}$$

Now we consider the Fitingof code applied to n-letter sequences generated by a Markov chain μ of memory m over some alphabet Λ. The Fitigof code is prefix-free and, hence, from (14) and (17) we obtain the following

Claim 3 *For any distribution μ*

$$h_{min}(\pi_{F,\mu}) > n^* - (|\Lambda|^m(|\Lambda| - 1)\log n\, +\, 2)\,. \tag{18}$$

In particular, for an i.i.d. source with binary alphabet

$$h_{min}(\pi_{F,\mu}) > n^* - (\log n + 2)\,.$$

5.3 Description of the cipher for Markov chains

Now we can apply the cipher $cipher_{rw-ds}$ from the part 3 with $\phi = \phi_F$ and $\pi = \pi_{F,\mu}$. (Recall that this cipher is one for all distributions with equal min-entropy and, hence, it does not depend on unknown distribution μ.)

From (18) we can see that $\Delta = |\Lambda|^m(|\Lambda| - 1)\log n\, +\, 2$ and applying theorem 3 and 4 and this estimate we obtain

Claim 4 *Let there be set of messages from Λ^n, $n > 0$, generated by Markov chain of order m, $m \geq 0$, and $\epsilon > 0$. Let the cipher $cipher_{c\&r}$ be applied.*

(i) *If the key length k equals $(|\Lambda|^m(|\Lambda| - 1)\log n\, +\, 5\, + 2\log(1/\epsilon) +$ bits, then the cipher is ϵ- entropically secure.*

(ii) *If the key length k equals $k = (|\Lambda|^m(|\Lambda| - 1)\log n\, +\, 9)\, +\, 2log(1/\epsilon)$ bits, the cipher is ϵ- indistinguishable.*

Data availability All data generated or analysed during this study are included in this published article.

Appendix

The definition of a stationary ergodic Markov chain with memory, or connection, *m*

First we give a definition of stationary ergodic processes. The time shift T on Λ^∞ is defined as $T(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots)$. A process P is called stationary if it is T-invariant: $P(T^{-1}B) = P(B)$ for every Borel set $B \subset \Lambda^\infty$. A stationary process is called ergodic if every T-invariant set has probability 0 or 1: $P(B) = 0$ or 1 whenever $T^{-1}B = B$ [2, 19].

We denote by $M_\infty(\Lambda)$ the set of all stationary and ergodic sources and let $M_0(\Lambda) \subset M_\infty(\Lambda)$ be the set of all i.i.d. processes. We denote by $M_m(\Lambda) \subset M_\infty(\Lambda)$ the set of Markov sources of order (or with memory, or connectivity) not larger than m, $m \geq 0$. By definition $\mu \in M_m(\Lambda)$ if

$$\mu(x_{t+1} = a_{i_1} | x_t = a_{i_2}, x_{t-1} = a_{i_3}, \ldots, x_{t-m+1} = a_{i_{m+1}}, \ldots) \\ = \mu(x_{t+1} = a_{i_1} | x_t = a_{i_2}, x_{t-1} = a_{i_3}, \ldots, x_{t-m+1} = a_{i_{m+1}})$$

for all $t \geq m$ and $a_{i_1}, a_{i_2}, \ldots \in \Lambda$.

Entropically secure ciphers

In this part we describe one entropically secure cipher from [5], part 3.2.

Let $\{h_i\}_{i \in I}$ be some family of functions $h_i : \{0, 1\}^k \to \{0, 1\}^n$, indexed over the set $I = \{0, 1\}^r$. By definition, a collection of functions from n-bit words to n-bits is XOR-universal if:

$$\forall a, x, y \in \{0, 1\}^n, x \neq y, Pr\{h_i(x) \oplus h_i(y) = a\} \leq \frac{1}{2^{n-1}},$$

if i is randomly chosen from I according to the uniform distribution ($\oplus$ is symbol-by-symbol modulo 2 summation). Also, suppose that there is a XOR-universal collection of functions whose description is public and, hence, it is known to Alice, Bob and Eve.

Dodis and Smith consider an encryption scheme of the form

$$E(m, K, i) = (i; m \oplus h_i(K))$$

where i is randomly chosen from I according to the uniform distribution, and K is a k-bit secret key. Note that m is a ciphered message of length n, i is the number of h_i in the set I and $|i| = \log |I| = r$. (Dodis and Smith notice that this scheme is a special low-entropy, probabilistic one-time pad). Decryption is obviously possible, since the description of the function h_i is public. It is shown [5] that this cipher is ϵ-entropically secure for $|k| \geq n - h_{min} + 2\log(1/\epsilon) + 2$ if the function family $\{h_i\}_{i \in I}$ is XOR-universal.

An example of XOR-universal family is as follows [5]: View $\{0, 1\}^n$ as $\mathcal{F} = GF(2^n)$, and embed the key set $\{0, 1\}^k$ as a subset of $\mathcal{F}$. For any $i \in \mathcal{F}$, let $h_i(K) = iK$, with multiplication in $\mathcal{F}$. This yields a family of linear maps $\{h_i\}$ with 2^n members. For this family the complexity of ciphering and deciphering is $O(n \log n \log \log n)$ [5].

It is important to note that the length of the secret key (k) depends only on the min-entropy of the probability distribution and does not depend on other parameters of the distribution.

References

1. Agrikola T., Couteau G., Ishai Y., Jarecki S., Sahai A.: On pseudorandom encodings. In: Theory of Cryptography Conference, pp. 639–669. Springer, Cham (2020).
2. Billingsley P.: Ergodic Theory and Information. Wiley, Hoboken (1965).
3. Calmon F.D.: Information-theoretic metrics for security and privacy (Doctoral dissertation, Massachusetts Institute of Technology) (2015).
4. Cover T.M., Thomas J.A.: Elements of information theory. Wiley, New York (2006).
5. Dodis Y., Smith A.: Entropic security and the encryption of high entropy messages. In: Theory of Cryptography Conference, pp. 556–577. Springer, Berlin (2005).
6. du Pin Calmon F., Medard M.L.M., Zeger L.M., Barros J., Christiansen M.M., Duffy K.R.: 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, pp. 1387–1394. IEEE (2012).
7. Elias P.: The efficient construction of an unbiased random sequence. Ann. Math. Stat. **43**(3), 864–870 (1972).
8. Fitingof B.M.: Optimal coding in the case of unknown and changing message statistics. Probl. Peredachi Inform. **2**(2), 3–11 (1966).
9. Gilbert E.W., Moore E.F.: Variable length binary encoding. Bell. Syst. Tech. J. **38**, 933–967 (1959).

10. Gunther C.G.: A universal algorithm for homophonic coding. In: Workshop on the Theory and Application of Cryptographic Techniques, pp. 405–414. Springer, Berlin (1988).
11. Jaeger J., Ristenpart T., Tang Q.: Honey encryption beyond message recovery security. In: Advances in Cryptology-EUROCRYPT 2016: 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8–12, Proceedings, Part I 3, pp. 758–788. Springer, Berlin (2016).
12. Juels A., Ristenpart T.: Honey encryption: security beyond the bruteforce bound. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques, pp. 293–310. Springer, Berlin (2014).
13. Krichevsky R.: Universal compression and retrival. Kluver Academic Publishers, New York (1993).
14. 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). Schloss Dagstuhl-Leibniz-Zentrum Informatik (2021).
15. Russell A., Wang H.: How to fool an unbounded adversary with a short key. IEEE Trans. Inf. Theory. **52**(3), 1130–40 (2006).
16. Ryabko B.Y.: The fast enumeration of combinatorial objects. Discret. Math. Appl. **10**(2), 163–182 (1998).
17. Ryabko B.: A simply realizable ideal cryptographic system. Probl. Inf. Transm. **36**(1), 84–89 (2000) **(see also IACR Cryptology ePrint archive, report 2001/046)**.
18. Ryabko B.: The Vernam Cipher is robust to small deviations from randomness. Probl. Inf. Transm. **51**(1), 82–86 (2015).
19. Ryabko D.: Asymptotic nonparametric statistical analysis of stationary time series. Springer, New York (2019).
20. Ryabko B., Fionov A.: Efficient homophonic coding. IEEE Trans. Inf. Theory **45**(6), 2083–2091 (1999).
21. Ryabko B., Matchikina E.: Fast and efficient construction of an unbiased random sequence. IEEE Trans. Inf. Theory **46**(3), 1090–1093 (2000).
22. Ryabko B., Ryabko D.: Information–theoretic approach to steganographic systems. In: IEEE International Symposium on Information Theory, Proceedings, 2461–2464. https://eprint.iacr.org/2006/063 (2007).
23. Shannon C.E.: Communication theory of secrecy systems. Bell Syst. Tech. J. **28**(4), 656–715 (1949).
24. von Neumann J.: Various techniques used in connection with random digits. Nat. Bur. Stand. Appl. Math. Ser. **12**, 36–38 (1951) **(Reprinted in the Collected Works of von Neumann, vol. 5)**.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.