Using data compression and randomisation to build an unconditionally secure short key cipher

Boris Ryabko
Federal Research Center for Information and Computational Technologies of SB RAS
Novosibirsk state university
Novosibirsk, Russian Federation
Email: boris@ryabko.net

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 data compression and randomisation techniques combined with entropically-secure encryption for the case when message statistics are known. The resulting cipher can be used for encryption in such a way that the key length does not depend on the entropy or the length of the encrypted message; instead, it is determined by the required security level.

keywords: cryptography, unconditionally secure cipher, entropically-secure encryption, Shannon code, randomisation,

I. INTRODUCTION

The concept of unconditional secrecy was presented in the seminal article by C. Shannon, where he also showed that a one-time pad (or Vernam cipher) is unconditionally secure [1]. 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 approaches for creating secure ciphers with short or low entropy keys, see [2]–[11]. One such approach was suggested by C. Shannon in [1], 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 [6].

A promising new way of building secure systems with short keys is the so-called entropic security, proposed by Russell and Wang [2] and developed by Dodis and Smith [3]. 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 [6], [9].

The notion of an entropy-secure symmetric encryption scheme is extremely important for cryptography because one can construct entropically-secure symmetric encryption schemes with keys much shorter than the length of the input, thus circumventing Shannon’s famous lower bound on key length [3].

Another way to construct a short key cipher is the so-called honey cipher, proposed by Jewels and Ristenpart [7] and developed by Joseph, Ristenpart and Tang [8]. In some ways, the honey cipher is like the ideal cipher: a computationally unlimited adversary has many possible highly probable decryptions. Li, Tang and Zhang [12] developed and combined the ideas of the honey cipher and the entropy 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 [1], [13]. Moreover, homophonic coding can be used to compress and randomise together [13], [14]. The goal of both transformations is to make the probability distribution of the original messages closer to the uniform one (see an overview in [15]). Interestingly, both transformations have been successfully applied to some cryptographic problems: they were used to extract randomness [16]–[18] and to build an ideal steganographic system [19].

In this article, we combine entropically-secure encryption with the suggested compression and data randomization techniques. Using compression and randomization, the original message is transformed in such a way that the difference between its length and its min-entropy is constant. 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 (but depends on the required security level).

II. 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 know $p$ and have a shared secret key $K = K_1...K_n$, which can be much shorter than the length...
of $M$, that is $k < n$. Alice encrypts $M$ with $K$ and possibly some random bits, and obtains the word $\text{cipher}(M, K)$. 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.

Russell and Wang [2] suggested a definition of the entropic security which was generalised by Dodis and Smith [3] as follows: A probabilistic map $Y$ is said to hide all functions on $\{0, 1\}^n$ with leakage $\epsilon$ if, for every adversary $A$, there exists some random bits, and obtains the word $\text{cipher}(M, K)$.

Definition 1: The map $Y()$ is called $(t, \epsilon)$-entropically secure if $Y()$ hides all functions on $\{0, 1\}^n$, whenever the min-entropy of the probability distribution $p$ is at least $t$, where min-entropy $h_{\min}(p)$ is as follows:

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

(2)

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

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

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

(3)

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

The following theorem of Dodis and Smith [3] is a generalisation of the results of Russell and Wang [2].

Theorem 1: ( [3]) Let there be an alphabet $\{0, 1\}^n, n > 0$, with a probability distribution whose min-entropy equals $h_{\min}$. Then there exists an efficient $(h_{\min}, \epsilon)$-entropically secure cipher with the $k$-bit key where

$$k = n - h_{\min} + 2\log(1/\epsilon) + 2 .$$

(4)

Another important notion is that of indistinguishability:

Definition 2: ( [3]). A randomized map $Y()$ is $(t, \epsilon)$-indistinguishable if there is a random variable $G$ such that for every distribution on messages $M$ over $\{0, 1\}^n$ with min-entropy at least $t$, we have

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

where for two distributions $A, B$

$$SD(A, B) = \frac{1}{2} \sum_{M \in M} |\Pr\{A = M\} - \Pr\{B = M\}| .$$
Also, it is known that the set of codewords $\lambda(a_1), ..., \lambda(a_L)$ 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 $v$ or $v$ is a prefix of $u$.)

Note that, for any sequence $x_1x_2...x_n$, $n \geq 1$, from the alphabet $A$ and a prefix-free code $\lambda$ the encoded sequence $\lambda(x_1)\lambda(x_2)\lambda(x_n)$ can be decoded to $x_1x_2...x_n$ without errors. Such a code $\lambda$ 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 $\lambda$, that is, it is the prefix-free and (5) is valid. (That is why we do not describe the transformation of $\hat{\lambda}$ and $\lambda$ in detail and do not estimate its complexity.)

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

In the following 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 $\lambda$ is a code then

$$\lambda^t(a_i) = \begin{cases} 0 & \text{if } |\lambda(a_i)| \leq \lceil \log L \rceil \\
1 \text{ bin}_{\log L}(i) & \text{if } |\lambda(a_i)| > \lceil \log L \rceil ,
\end{cases}$$

where $\text{bin}_{\log L}(i)$ is a binary presentation of $i$ whose length is $\lceil \log L \rceil$. (For example, $\text{bin}_{3}(3) = 011$). 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 $\lambda$ 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 (6) and (5) we obtain

$$|\lambda^t_{Sh}(a_i)| \leq \lfloor \log(1/p(a_i)) \rfloor + 1 < \log(1/p(a_i)) + 2. \quad (7)$$

B. Randomised prefix-free codes

Let $\lambda$ be a prefix-free code for the alphabet $A$ and

$$l = \max_{i=1,...,L} |\lambda(a_i)|.$$

The randomized code $\rho_\lambda$ maps letters from the alphabet $A$ to the set $\{0, 1\}^l$ defined as follows.

$$\rho_\lambda(a_i) = \lambda(a_i) r_1^{\lceil \log(1/p(a_i)) \rceil + 1} r_2^{\lceil \log(1/p(a_i)) \rceil + 2} ... r_l^1,$$  \quad (8)

where $r_1^{\lceil \log(1/p(a_i)) \rceil + 1}, r_2^{\lceil \log(1/p(a_i)) \rceil + 2} ... r_l^1$ uniformly distributed and independent random bits (for all $i$). Let us define a probability distribution $\pi_\lambda$ on $\{0, 1\}^l$ as follows:

$$\pi_\lambda(y_1y_2...y_l) = p(a_i)2^{-l(|\lambda(a_i)|)}$$

if $y_1y_2...y_l = \lambda(a_i). \quad (9)$

If for some $y = y_1...y_l$ any $\lambda(a_i)$ is not a prefix of $y$, then $\pi_\lambda(y) = 0$.

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

$$h_{min}(\pi_{\lambda_{Sh}}) > l - 1, \quad h_{min}(\pi_{\lambda_{Sh}}) > l - 2. \quad (10)$$

Here the first equation follows from the definition of the min-entropy and (III-B), whereas (10) follows from (5) and (7).

C. The cipher

Here we describe a cipher with the length of key $2\log(1/\epsilon) + O(1)$ which is $(0, \epsilon)$ entropically secure. It means that this cipher is $\epsilon$ entropically secure for any probability distribution (i.e. with any min-entropy). This cipher is based on the application of the entropically secure cipher of Dodis and Smith [3] to the suggested randomized code. We describe the suggested cipher for the $\lambda_{Sh}^t$ code, but expanding to other prefix-free codes is straightforward.

The suggested ciphering is carried out in the following three steps: for a given $\epsilon \in (0, 1)$

i) Build a code $\lambda^t_{Sh}$ for the alphabet $\{0, 1\}^n$ with a probability distribution $p$.

ii) Calculate $l = \max_{a \in \{0, 1\}^n} |\lambda^t_{Sh}(a)|$ and probabilities $\pi_{\lambda_{Sh}}(u), u \in \{0, 1\}^l$. (From (6) we can see that $l \leq n + 1$.)

iii) For the alphabet $\{0, 1\}^l$ with the distribution $\pi$ build $(l - 2, \epsilon)$ entropically secure cipher of Dodis and Smith [3] with the $k$-bit key where

$$k = 2\log(1/\epsilon) + 4.$$

Note that Dodis and Smith [3] proposed three entropically secure ciphers, any of which can be used; so we do not describe any particular cipher here.

From the Dodis and Smith Theorem 2 (see (3)) and the estimate of the min-entropy (10) we can see that such a cipher exists for the distribution $\pi_\lambda$ on $\{0, 1\}^l$.

Strictly speaking, we are dealing with two ciphers. One is applied to the set $\{\rho_\lambda(M) \in \{0, 1\}^l\}$, whereas the second (the main one) is applied to $\{M \in \{0, 1\}^n\}$, and include data compression and randomisation. We denote the first one cipher$_{ds}$ and the second cipher$_{ckr}$.

The Theorem 2 guarantees the entropic security for the first cipher cipher$_{ds}$, so, we need to prove a similar property for cipher$_{ckr}$. The following theorem describes the properties of this cipher:

Theorem 3: Let any message $M$ belong to $\{0, 1\}^n, n > 0$, and they obey some probability distribution $p$. Let $\epsilon > 0$. Then, cipher$_{ckr}$ is $(0, \epsilon)$ entropically secure with the $k$-bit key with
Let $k = 2\log(1/\epsilon) + 4$, that is, for any function $A : \{0,1\}^l \rightarrow \{0,1\}^n$ and $f : \{0,1\}^n \rightarrow \{0,1\}^*$

$$|Pr\{A(\text{cipher}_{c\kappa r}(M)) = f(M)\} - Pr\{\hat{A}(\cdot) = f(M)\}| \leq \epsilon,$$

where $\hat{A}$ does not use $\text{cipher}_{c\kappa r}(M)$.

Comment. This means that the cipher is $\epsilon$-entropically secure with the secret keys whose lengths do not depend on the probability distribution on the set of the messages $M \subset \{0,1\}^n$.

Proof. From Theorem 1 we see that

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

where $v$ is random variable with distribution $\{\pi(x), M \in \{0,1\}^n\}$, $g$ is any function defined on $\{0,1\}^l (g : \{0,1\}^l \rightarrow \{0,1\}^*)$ and $\hat{A}(\cdot)$ does not depend on $v$ (to be short, $\lambda = \lambda_{Sh}$).

Taking into account that the code $\lambda$ is prefix-free, we can define such a function $\phi$ that for any $a \in \{0,1\}^*$ and $u = \rho_\lambda(a)$, $\phi(u) = a$. For any function $f : \{0,1\}^n \rightarrow \{0,1\}^*$ and $M$ consider the function $g(\rho_\lambda(M)) = f(\phi(\rho_\lambda(M))) = f(M)$. The last inequality is valid for this function $g$ and for $v = \rho_\lambda(M)$, hence

$$|Pr\{A(cipher_{ds}(\rho_\lambda(M))) = f(\phi(\rho_\lambda(M)))\} - Pr\{\hat{A}(\cdot) = f(\phi(\rho_\lambda(M)))\}| \leq \epsilon.$$

Taking into account that $cipher_{c\kappa r}(M) = cipher_{ds}(\rho_\lambda(M))$ and $f(\phi(\rho_\lambda(M))) = f(M)$, we can see from the latest inequality that

$$|Pr\{A(cipher_{c\kappa r}(M)) = f(M)\} - Pr\{\hat{A}(\cdot) = f(M)\}| \leq \epsilon.$$

The theorem is proven.

We can see from Theorem 2 that the cipher $cipher_{c\kappa r}$ is $2\epsilon$ indistinguishable. Below we directly prove that $cipher_{c\kappa r}$ is $\epsilon$ indistinguishable with $k = 2\log(1/\epsilon) + 5$.

Theorem 4: Let there be an alphabet $\{0,1\}^n, n > 0$, with some probability distribution. If, instead of (11), the key length $k$ equals $k = 2\log(1/\epsilon) + 5$, the cipher is $(0, \epsilon)$ indistinguishable.

Proof. From Theorem 2 we know that, in fact, the indistinguishability is equal to the entropic security, and it is valid for $cipher_{ds}$, but we are interested in the indistinguishability of the $cipher_{c\kappa r}$. In order to prove it suppose that $cipher_{ds}$ is applied to the words from the set $\rho_\lambda(M) \subset \{0,1\}^l$ in such a way that it is $\epsilon, (\epsilon/4)$ entropically secure, where the length of the key equals $2\log(1/(\epsilon/4)) + 1 = 2\log(1/\epsilon) + 5$. From Theorem 2 we can see that this cipher is $(2, \epsilon)$ indistinguishable, that is, $SD(cipher_{ds}, G) \leq \epsilon$, where $G$ is a random variable on $\{0,1\}^l$ which is independent on $cipher_{ds}$.

Define $U_a = \{cipher_{ds}(\lambda(a) r) : r \in \{0,1\}^{l-\lambda(a)}\}$ and let $G'(v)$ be defined as follows:

$$Pr\{G' = v\} = \sum_{w \in U_v} Pr\{G = w\}.$$
[7] 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.

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

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

[10] Issa, I., Wagner, A. B., Kamath, S. An operational approach to information leakage. IEEE Transactions on Information Theory, 2019, 66(3), 1625-1657

[11] Shikel, Y. Y., Blum, R. S., Poor, H. V. (2020). Secrecy by design with applications to privacy and compression. IEEE Transactions on Information Theory, 2020, 67(2), 824-843.

[12] 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 für Informatik.

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

[14] Ryabko, B. Fionov, A. Efficient homophonic coding. IEEE Trans. Inform. Theory 45 (1999), no. 6, 2083–2091.

[15] Agrikola T., Couteau G., Ishai Y., Jarecki S., Sahai A. On pseudorandom encodings. In Theory of Cryptography Conference 2020 Nov 16 (pp. 639-669). Springer, Cham.

[16] J. von Neumann, Various techniques used in connection with random digits, Nat. Bur. Stand. Appl. Math. Ser., vol. 12, pp. 36738, 1951 (Reprinted in the Collected Works of von Neumann, vol. 5).

[17] P. Elias, The efficient construction of an unbiased random sequence, Ann. Math. Statist., vol. 43, no. 3, pp. 864/870, 1972.

[18] Ryabko, B.; Matchikina, E. Fast and efficient construction of an unbiased random sequence. IEEE Transactions on Information Theory, V.46, n.3, 2000, p.1090 -1093.

[19] Ryabko B. , Ryabko D. Information–Theoretic Approach to Steganographic Systems. IEEE International Symposium on Information Theory, Proceedings. 2007, Nice, France, 2461-2464. (see also https://eprint.iacr.org/2006/063).

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

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

[22] Ryabko B.Ya. The fast enumeration of combinatorial objects, Discrete Math and Applications, v.10, n. 2, 1998, 163-182

[23] Billingsley P. Ergodic Theory and Information,1965, John Wiley and Sons, Inc.

[24] Ryabko D. Asymptotic nonparametric statistical analysis of stationary time series. Springer, 2019