Employment of Homophonic Coding for Improvement of Certain Encryption Approaches Based on the LPN Problem

Miodrag Mihaljević* and Hideki Imai**

Abstract. This paper proposes an improvement of certain encryption approaches designed based on hardness of the learning from parity with noise (LPN) problem. The proposal employs a dedicated homophonic coding and randomness resulting in a harder underlying LPN problem in comparison with the related source schemes without homophonic coding. It is shown that the proposed scheme provides the following security features: (i) the indistinguishability in the CPA scenario; and (ii) hardness of the algebraic recovering of the secret key in the CPA scenario. Regarding (ii) it is shown that the secret key recovery is as hard as the LPN problem where the noise is equal to $\frac{1-(1-2p)^{(m-\ell)/2}}{2}$ and m, ℓ and p are the parameters of the proposed scheme. Consideration of the implementation complexity shows that it is low (regarding the both: time and space), assuming that the appropriate efficient linear block codes are employed. The proposed encryption is compared with the related recently reported ones and it is pointed out that the novel scheme can provide an enhanced security, reduced communications overhead and has approximately the same implementation complexity.

Keywords: symmetric encryption, LPN problem, randomness, homophonic coding, error-correction coding.

1 Introduction

Usefulness of involvement pure randomness into a cryptographic primitive has been recognized in a number of reported designs and particularly in the following ones. In [16], a number of approaches for including randomness in the encryption techniques have been discussed mainly regarding block and stream ciphers. According to [16], the randomized encryption is a procedure which enciphers a message by randomly choosing a ciphertext from a set of ciphertexts corresponding to the message under the current encryption key, and the following is claimed, [16]: "At the cost of increasing the required bandwidth, randomized encryption procedures may achieve greater cryptographic security than their deterministic counterparts ...". In [3], a pseudorandom number generator based on the Learning from Parity with Noise (LPN) problem, derived from an older proposal of one-way function based on the hardness of decoding a random linear code, has been reported. (Informally note that the LPN problem can be considered as the problem of solving a system of linear equations corrupted by noise. or a problem of decoding a linear code; A more formal specification of the LPN problem will be given later on). In [7], a probabilistic private-key encryption scheme named LPN-C whose security can be reduced to the hardness of the LPN problem has been proposed and considered. Recently, in [1] a symmetric encryption scheme similar to the one reported in [7] is reported and its security and implementation complexity are analyzed. The symmetric encryption schemes reported in [7] and [1] appears as interesting and stimulating for further considerations (having in mind improvements as well) particularly because the security is related to the recognized hard (LPN) problem.

* Mathematical Institute, Serbian Academy of Sci. and Arts, Kneza Mihaila 36, Belgrade, Serbia, and Research Center for Information Security (RCIS), National Institute of Advanced Industrial Science and Technology (AIST), Room 1003, Akihabara Daibiru, 1-18-13 Sotokanda, Chiyoda-ku, Tokyo, 101-0021 Japan. E-mail: miodragm@turing.mi.sanu.ac.rs .

** Faculty of Sciences and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551, Japan, and Research Center for Information Security (RCIS), National Institute of Advanced Industrial Science and Technology (AIST), Room 1003, Akihabara Daibiru, 1-18-13 Sotokanda, Chiyoda-ku, Tokyo, 101-0021 Japan.

Origins for the construction given in this paper are the approaches for stream ciphers design recently reported in [14] and [15] but the construction in this paper is substantially different because it does not employ the keystream generators. This difference has a number of implications regarding the security analysis and implementation complexity of the scheme. Motivation for this paper is a consideration of the possibilities for some novel approaches for inclusion of pure randomness into certain encryption framework. Particular goals of the paper are the following: (i) specification of a family of ciphers which involve (balanced) randomness and dedicated coding; (ii) consideration of the impact of randomness on the security of the proposed class of ciphers and the security statements based on the LPN problem hardness; (iii) a discussion on the implementation complexity and the communications overhead of the proposed class of ciphers; (iv) a comparison of the proposed ciphers family with the related recently reported encryption schemes.

Brief Summary of the Results.
This paper proposes an improvement of certain encryption approaches based on the LPN problem reported in [7] and [1]. The improvement is based on involvement of a dedicated homophonic coding with impact on the parameters of the LPN problem required for achieving certain security level, i.e. the involved homophonic encoding transforms an LPN type problem with a certain noise rate into another LPN type problem with a larger noise rate, thereby making the problem harder[1]. The proposed framework provides enhanced security and reduction of the communication overhead and low implementation complexity approximately same as the implementation complexity of the source scheme. The analysis shows that the security enhancement appears as a joint effect of pure randomness involved in homophonic encoding and the secret key in the following manner: The pure randomness involved in homophonic encoding additionally "protects" the secret key, and the secret key makes homophonic decoding hard for an attacker who does not posses the secret key, i.e. removing of the randomness appears as hard as recovering of the secret key. Regarding implementation complexity note that the involvement of linear homophonic coding implies that the encoding and decoding operations are vector-matrix multiplications. Also, note that assuming linear block error-correction coding, the homophonic and error-correction encoding can be considered as an equivalent encoding performed via single vector-matrix multiplication. Finally, note that the approaches [7] and [1] require a source of randomness and this source could be employed for generation of the randomness required for homophonic encoding as well. The proposed encryption is compared with the related ones reported in [7] and [1] and it is pointed out that increased security and reduced communication overhead can be achieved preserving the implementation complexity at the same time. Accordingly, the approach reported in this paper provides a significant improvement of the related previously reported ones, and includes a substantival novelty regarding a dedicated involvement of the homophonic coding for the security enhancement.

Organization of the Paper. Section 2 yields a summary of the backgrounds. Section 3 proposes an improved ciphering approach based on employment of homophonic coding including underlying ideas for the design and its algebraic structure. Security evaluation of the proposal is given in Section 4. Implementation complexity and communications overhead of the proposed ciphering

[1] The security evaluation implies that the homophonic encoding transforms the underlying LPN problem into another one where the noise level is increased. Accordingly, the following question could be raised: If the underlying problems are the same up to the nose parameter why not just propose the scheme with an increased noise. Regarding the addressed issue it is important to note that the homophonic encoding plays role of an amplifier of the initial noise related to the LPN problem only when the secret key is not known. Accordingly, a legitimate user does not face the increased noise but an attacker faces the LPN problem with increased noise. On the other hand, the employed level of the initial noise implies the required error-correction code and if this noise is high the required error-correction coding introduces a high (and unacceptable) overhead to the system.

approach are considered in Section 5. A comparison of the proposed construction with two related and recently reported ones is given in Section 6. Certain concluding notes are pointed out in the Section 7.

2 Background

2.1 The LPN Problem

The security of the encryption schemes which are origins for the one proposed in this paper, as well as the security of the proposed scheme, is related to the following informally specified problem (which is a hard one). Let k be a security parameter. If $\mathbf{s}, \mathbf{d}_1, ..., \mathbf{d}_q$ are binary vectors of length k, let $y_i =< \mathbf{s} \cdot \mathbf{d}_i >$ denote the inner product of $\mathbf{s}$ and $\mathbf{d}_i$ (modulo 2). Given the pairs $(\mathbf{d}_1, y_1)$, $(\mathbf{d}_2, y_2)$, ..., $(\mathbf{d}_q, y_q)$, for randomly-chosen $\{\mathbf{d}_i\}_{i=1}^q$ and $q = O(k)$, it is possible to efficiently determine $\mathbf{s}$ using standard linear-algebraic techniques. However, in the presence of noise where each y_i is flipped (independently) with probability ϵ, finding $\mathbf{s}$ becomes much more difficult. We refer to the problem of learning $\mathbf{s}$ in this case when the values $\{y_i\}_{i=1}^q$ are flipped as the learning parity in noise (LPN) problem with the parameters k, q and ϵ - $\text{LPN}_{k,q,\epsilon}$ problem (Formal definition of the LPN problem is out of the scope of this paper). Note that the LPN problem is a particular problem of solving a system of consistent overdefined linear equations corrupted with noise so that each equation is correct with a given probability.

2.2 Certain Encryption Schemes Based on the LPN Problem

Summary of the Symmetric Encryption Scheme [7] Let $C : \{0,1\}^\ell \rightarrow \{0,1\}^n$ be an $[n, \ell, d]$ error-correcting code (i.e. of length n, dimension ℓ, and minimal distance d) with correction capacity equal to the integer part of $t = \frac{d-1}{2}$. This error-correcting code is assumed to be publicly known. Let $\mathbf{S}$ be a secret key $k \times n$ matrix (constituting the secret key of the cryptosystem). To encrypt an ℓ-bit vector $\mathbf{a}$, the sender draws a k-bit random vector $\mathbf{u}$ and computes

$$\mathbf{z} = C(\mathbf{a}) \oplus \mathbf{u} \cdot \mathbf{S} \oplus \mathbf{v}, \tag{1}$$

where $\mathbf{v} \leftarrow \text{Ber}_{n,p}$ is an n-bit noise vector such that each of its bits is (independently) 1 with probability p and 0 with probability $1 - p$. The resulting "ciphertext" is the pair $(\mathbf{u}, \mathbf{z})$. Upon reception of this pair, the receiver decrypts by computing $\mathbf{z} \oplus \mathbf{uS} = C(\mathbf{a}) \oplus \mathbf{v}$, and decoding the resulting value. If decoding is not possible (which may happen when the code is not perfect), then the decryption algorithm returns "decryption error". When the message is not ℓ-bit long, it is padded till its length is the next multiple of ℓ and encrypted block-wise. Finally note that when the employed error-correcting code is a binary linear one, $C(\mathbf{a}) = \mathbf{a} \cdot \mathbf{G}$ where $\mathbf{G}$ is a binary matrix of dimension $\ell \times n$.

Summary of the Symmetric Encryption Scheme [1] The symmetric-key cryptosystem prposed in [1] is based on the LPN problem. Its ciphertexts are only a constant factor larger than the plaintexts, and both encryption and decryption can be performed by Boolean circuits of (approximately) linear size (in the message length), which is almost optimal even for standard CPA-security. The scheme is a close variant of the LPN-based encryption scheme reported in [7], which was proved secure only in the standard sense (i.e., without key-dependent messages), and did not achieve linear time efficiency. The symmetric encryption scheme proposed in [1] is summarized by the following.

Let $\ell = \ell(k)$ be a message-length parameter which is set to be an arbitrary polynomial in the security parameter k assuming that shorter messages are padded with zeroes, and let $\epsilon = 2^{-m}$

and $0 < \delta < 1$ be constants. The scheme [1] employs a family of good binary linear codes with information words of length $\ell(k)$ and block length $n = n(k)$, that has an efficient decoding algorithm D that can correct up to $(\epsilon+\delta)\cdot n$ errors. Accordingly, let $\mathbf{G} = \mathbf{G}_\ell$ be the $n \times \ell$ binary generator matrix of the employed code.

Let $N = N(k)$ be an arbitrary polynomial (which controls the tradeoff between the key-length and the time complexity of the scheme). The private key of the scheme is a matrix $\mathbf{S}$ which is chosen uniformly at random from $\mathcal{Z}_2^{k\times N}$ Then, the encryption and decryption procedures are specified as follows.

Encryption: To encrypt a message in form of a matrix $\mathbf{A} \in \mathcal{Z}_2^{\ell\times N}$, choose a balanced random matrix $\mathbf{U} \leftarrow \mathcal{Z}_2^{n\times k}$ and a random noise matrix $\mathbf{V} \leftarrow \mathrm{Ber}_\epsilon^{n\times N}$. Output the ciphertext $(\mathbf{U}, \mathbf{Z})$, where

$$\mathbf{Z} = \mathbf{G}\cdot\mathbf{A} \oplus \mathbf{U}\cdot\mathbf{S} \oplus \mathbf{V}\,. \tag{2}$$

Decryption: Given a ciphertext $(\mathbf{U}, \mathbf{Z})$ apply the decoding algorithm D to each of the columns of the matrix $\mathbf{Z} \oplus \mathbf{U}\cdot\mathbf{S}$ and output the result.

Observe that the decryption algorithm fails only when there exists a column in $\mathbf{V}$ whose Hamming weight is larger than $(\epsilon + \delta)n$.

2.3 Universal Homophonic Coding

Homophonic coding or "multiple substitution" (see [10], [17] and [12], for example) is a technique for mapping source data employing certain random bits into the encoded data which are the randomized form of the source ones so that the source data can be recovered from the noise-free encoded ones without knowledge of the random bits. Homophonic encoding provides that many particular outputs of encoding become possible substitutes (or "homophones") of the source data based on employment of different random sequences. Perfect homophonic code provides that the encoded data appear as truly random ones.

A particular class of homophonic codes are the universal ones reported in [12]: These codes provide the randomization without knowledge of the source data statistics which is a request for some homophonic coding schemes. The source data can be recovered from the homophonic encoder output without knowledge of the randomizing data by passing the encoded data through the decoder and then discarding the randomizer bits.

Finally note that the Wire-tap channel coding [19] is based on assigning multiple codewords to the same information vector and from that point of view, particularly when the main channel is noise-free, it shares the same underlying idea employed in the homophonic coding. (In the following, only the homophonic coding will be considered.)

3 An Encryption Scheme Based on Homophonic Coding and its Algebraic Representation

3.1 Underlying Ideas for Improvement the Reported Approach

It is well known (see [16], for example) that involvement of randomness in a cryptographic primitive can result into an enhanced security. Let $\mathbf{a}$ be a binary vector which is subject of homophonic encoding and $\mathbf{r}$ be a random binary vector. A linear homophonic encoding performs mapping of the concatenation of $\mathbf{a}$ and $\mathbf{r}$ into a resulting binary vector, a codeword, where each bit is a linear combination of certain bits of $\mathbf{a}$ and $\mathbf{r}$. If the codeword is encrypted by mod2 addition with certain secret binary vector, the resulting vector consists of bits which depend on certain random and certain secret data.

Our goal is to design an encryption scheme where, assuming the chosen plaintext attack, the randomness involved in homophonic encoding protects secret key as a consequence of the

following: Removing of the randomness, i.e. decoding, without knowledge of the secret key becomes as complex as recovering the secret key employing the exhaustive search approach. (The security evaluation given in Section 4 shows how close the proposed design is to the above specified goal.)

Accordingly, this paper proposes employment of the concatenation of dedicated homophonic encoding and error-correction coding instead of just the error-correction one as the approach for enhancing the security, as well as to provide additional implementation flexibility of the encryption schemes reported in [7] and [1].

3.2 Proposal

The design proposed in this section originates from a consideration of the possibilities for some novel approaches regarding inclusion of pure randomness into the symmetric key encryption frameworks reported in [7] and [1]. The main goal of employment the pure randomness is to provide a supporting element for enhancing the security implied by hardness of the LPN problem.

We assume the following notations:
- $\mathbf{a} = [a_i]_{i=1}^{\ell}$: ℓ-dimensional binary vector of message/plaintext data;
- $\mathbf{r} = [r_i]_{i=1}^{m-\ell}$: $(m-\ell)$-dimensional binary vector of random data where each r_i is a realization of the binary random variable R_i such that $\Pr(R_i = 1) = \Pr(R_i = 0) = 1/2$, $i = 1, 2, ..., n$;
- $\mathbf{u} = [u_i]_{i=1}^{k}$: k-dimensional binary vector of random data where each u_i is a realization of the binary random variable U_i such that $\Pr(U_i = 1) = \Pr(U_i = 0) = 1/2$, $i = 1, 2, ..., k$;
- $\mathbf{S} = [s_{i,j}]_{i=1\ j=1}^{k\ n}$: $k \times n$-dimensional binary matrix of the secret key
- $\mathbf{v} = [v_i]_{i=1}^{n}$: n-dimensional binary vector of random data where each v_i is a realization of the binary random variable V_i such that $\Pr(V_i = 1) = p$ and $\Pr(V_i = 0) = 1 - p$, $i = 1, 2, ..., n$;
- $C_H(\cdot)$ and $C_H^{-1}(\cdot)$: operators of the homophonic / wire-tap channel encoding and decoding, respectively; $C_H(\cdot)$ denotes a mapping $\{0,1\}^m \rightarrow \{0,1\}^m$;
- $C_{ECC}(\cdot)$ and $C_{ECC}^{-1}(\cdot)$: operator of the error-correction encoding and decoding, respectively; $C_{ECC}(\cdot)$ denotes a mapping $\{0,1\}^m \rightarrow \{0,1\}^n$; .

This paper proposes a symmetric key encryption scheme where the encryption and decryption operations are specified by the following.

– *Encryption*
  1. Employing $\mathbf{r}$ perform the homophonic (wire-tap channel) encoding of the $\mathbf{a}$ and the error-correction encoding of the resulting vector as follows: $C_{ECC}(C_H(\mathbf{a}||\mathbf{r}))$ where $||$ denotes the concatenation.
  2. Generate the ciphertext in form of n dimensional binary vector $\mathbf{z}$ as follows:

$$\mathbf{z} = C_{ECC}(C_H(\mathbf{a}||\mathbf{r})) \oplus \mathbf{u} \cdot \mathbf{S} \oplus \mathbf{v} \ . \tag{3}$$

– *Decryption*
  Assuming availability of the pair $(\mathbf{u}, \mathbf{z})$ decrypt the ciphertext as follows:

$$\mathbf{a} = tcat_{\ell}(C_H^{-1}(C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S}))) \ , \tag{4}$$

  where $tcat_{\ell}(\cdot)$ denotes truncation of the argument vector to the first ℓ bits and the assumption is that the employed code which corresponds to $C_{ECC}(\cdot)$ and $C_{ECC}^{-1}(\cdot)$ can correct the errors introduced by a binary symmetric channel with the crossover probability p.

Note that, as in the case of the ciphering schemes [7] and [1], the random vector $\mathbf{u}$ is a public one, and the decryption part assumes availability of the pair ($\mathbf{u}$, *ciphertext*). Also note that the decryption does not require knowledge of $\mathbf{r}$.

3.3 Algebraic Structure Assuming Employment of Linear Codes

Encoding and Decoding *Encoding Issues.* When the employed homophonic and error-correcting codes are linear, the encoding operations in the both cases are vector-matrix multiplications. Accordingly, the encoded version of $\mathbf{a}||\mathbf{r}$ is given by the following:

$$C_H(\mathbf{a}||\mathbf{r}) = [\mathbf{a}||\mathbf{r}]\mathbf{G}_H, \tag{5}$$

and $\mathbf{G}_H$ is an $m \times m$ matrix, and thus

$$\begin{aligned} C_{ECC}(C_H(\mathbf{a}||\mathbf{r})) &= C_{ECC}([\mathbf{a}||\mathbf{r}]\mathbf{G}_H) \\ &= [\mathbf{a}||\mathbf{r}]\mathbf{G}_H\mathbf{G}_{ECC} \\ &= [\mathbf{a}||\mathbf{r}]\mathbf{G} \end{aligned} \tag{6}$$

where $\mathbf{G}_{ECC}$ is an $m \times n$ binary generator matrix corresponding to $C_{ECC}(\cdot)$, and $\mathbf{G} = \mathbf{G}_H\mathbf{G}_{ECC}$ is an $m \times n$ binary matrix summarizing the two successive encodings at the encryption side, implying that

$$\mathbf{z} = [\mathbf{a}||\mathbf{r}]\mathbf{G} \oplus \mathbf{u} \cdot \mathbf{S} \oplus \mathbf{v} \ .$$

Decoding Issues. Assuming that the employed ECC can correct all the errors introduced by the vector $\mathbf{v}$ we have

$$C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S}) = C_{ECC}^{-1}(C_{ECC}(C_H(\mathbf{a}||\mathbf{r})) \oplus \mathbf{v}) = [\mathbf{a}||\mathbf{r}]\mathbf{G}_H \ , \tag{7}$$

and accordingly,

$$C_H^{-1}(C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S})) = [\mathbf{a}||\mathbf{r}]\mathbf{G}_H\mathbf{G}_H^{=1} = [\mathbf{a}||\mathbf{r}] \ , \tag{8}$$

implying that

$$tcat_\ell(C_H^{-1}(C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S}))) = \mathbf{a} \ .$$

An Algebraic Representation at the Bit Level Let $\mathbf{G} = [g_{i,j}]_{i=1}^{m}{}_{j=1}^{n}$, and let $\mathbf{z} = [z_i]_{i=1}^{n}$. Then,

$$z_i = (\bigoplus_{j=1}^{\ell} g_{j,i}a_j) \oplus (\bigoplus_{j=1}^{m-\ell} g_{\ell+j,i}r_j) \oplus (\bigoplus_{j=1}^{k} s_{j,i}u_j) \oplus v_i \ , \ i = 1, 2, ..., n, \tag{9}$$

implying that under the known plaintext attack we have

$$x_i \oplus (\bigoplus_{j=1}^{m-\ell} g_{\ell+j,i}r_j) \oplus v_i = z_i \oplus (\bigoplus_{j=1}^{\ell} g_{j,i}a_j) \ , \ i = 1, 2, ..., n, \tag{10}$$

where

$$x_i = (\bigoplus_{j=1}^{k} s_{j,i}u_j) \ , \ \ i = 1, 2, ..., n \ , \tag{11}$$

and where the right-hand side of the equations have known value.

In a particular setting, the homophonic encoding can be represented by the following matrix $\mathbf{G}_H$ assuming that it is the invertible one.

$$[\mathbf{a}||\mathbf{r}]\mathbf{G}_H = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_\ell \\ r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_{m-\ell} \end{bmatrix}^T \cdot \left[\begin{array}{cccc|ccc} g_{1,1}^{(H)} & g_{1,2}^{(H)} & \cdots & g_{1,m-\ell}^{(H)} & g_{1,m-\ell+1}^{(H)} & g_{1,m-\ell+2}^{(H)} & \cdots & g_{1,m}^{(H)} \\ g_{2,1}^{(H)} & g_{2,2}^{(H)} & \cdots & g_{2,m-\ell}^{(H)} & g_{2,m-\ell+1}^{(H)} & g_{2,m-\ell+2}^{(H)} & \cdots & g_{2,m}^{(H)} \\ \multicolumn{4}{c|}{\cdot} & & & & \\ \multicolumn{4}{c|}{\cdot} & & & & \\ \multicolumn{4}{c|}{\cdot} & & & & \\ g_{\ell,1}^{(H)} & g_{\ell,2}^{(H)} & \cdots & g_{\ell,m-\ell}^{(H)} & g_{\ell,m-\ell+1}^{(H)} & g_{\ell,m-\ell+2}^{(H)} & \cdots & g_{\ell,m}^{(H)} \\ \hline 1 & 0 & \cdots & 0 & g_{\ell+1,m-\ell+1}^{(H)} & g_{\ell+1,m-\ell+2}^{(H)} & \cdots & g_{\ell+1,m}^{(H)} \\ 0 & 1 & \cdots & 0 & g_{\ell+2,m-\ell+1}^{(H)} & g_{\ell+2,m-\ell+2}^{(H)} & \cdots & g_{\ell+2,m}^{(H)} \\ \multicolumn{4}{c|}{\cdot} & & & & \\ \multicolumn{4}{c|}{\cdot} & & & & \\ \multicolumn{4}{c|}{\cdot} & & & & \\ 0 & 0 & \cdots & 1 & g_{m,m-\ell+1}^{(H)} & g_{m,m-\ell+2}^{(H)} & \cdots & g_{m,m}^{(H)} \end{array}\right], \quad (12)$$

where the sub-matrix with the ones on the main diagonal and the zeros on all other positions is the square $(m-\ell)\times(m-\ell)$ one.

4 Security Evaluation of the Proposed Encryption Scheme

4.1 Attacking Scenario

Following the security evaluation claims given in [7] and [1], the security evaluation given in this section is related to the chosen plaintsxt attack (CPA) because the main goal of security evaluation is to point out the impacts of involvement of homophonic dedicated coding into the encryption framework proposed in [7] and [1].

Regarding the security of the encryption scheme proposed in this paper against the chosen ciphertext attacks (CCAs) when an attacker has access to the decryption oracle (including the adaptive CCA) note the following: (i) The proposed scheme is as vulnerable to a CCA as the ones reported in [7] and [1]; (ii) Security of the proposed scheme against CCAs can be achieved in the same manner as considered in [7] and [1] i.e. by using a Message Authentication Code (MAC). (For example the encrypt-then-MAC paradigm discussed in [7] can be employed.)

Accordingly, in order to point out the main effects of the involved homophonic coding, this section discusses the indistinguishability in CPA scenario, and security implied by hardness of recovering secret key based on the algebraic representation of encryption also in CPA scenario.

4.2 Security Implied by the Indistinguishability in CPA Scenario

The indistinguishability (IND) deals with the secrecy provided by the scheme in the following sense: An adversary must be unable to distinguish the encryption of two (chosen) plaintexts. Accordingly, and particularly following [7], as a criterion for the security consideration, in this section we consider the IND one.

Main claim of this section is that the proposed scheme fulfils the same IND security as the one in [7], i.e. that the employed homophonic coding does not affects the IND security, and so only a brief generic discussion is given.

For the IND considerations we assume the following traditional approach. An adversary is considered as a pair of algorithms $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ and they operate through two phases as follows.

- $\mathcal{A}_1$ is employed during the first phase and at the end of this phase, $\mathcal{A}_1$ outputs a pair of plaintexts $(\mathbf{a}_1, \mathbf{a}_2)$.

– One of the given plaintexts is selected with probability equal 1/2, then encrypted, and the obtained ciphertext is delivered to $\mathcal{A}_2$ - this represents $\mathcal{A}$'s challenge. The success of $\mathcal{A}$ is determined according to correctness of decision whether $\mathbf{a}_1$ or $\mathbf{a}_2$ was encrypted.

In [7] and [1] it is shown that the encryption schemes reported there (see Section 2.1, as well) are IND secure ones assuming that the related underlying LPN problems are hard. Consequently, we have the following statement.

Theorem 1. The encryption scheme proposed in this paper is IND secure assuming employment of a linear homophonic and error-correction coding.

Sketch of the Proof. Note that under the assumption of linearity, an error-correction encoding and a concatenation of homophonic and error-correction coding can be represented by a vector-matrix multiplication, i.e. all three schemes, the ones reported in [7] and the proposed one, have the following encryption framework

$$\mathbf{z} = C^*(\mathbf{a}) \oplus \mathbf{uS} \oplus \mathbf{v} \quad . \tag{13}$$

where the encoding operator $C^*(\cdot)$ represents either error-correction encoding or a concatenation of homophonic and error-correction encoding. Accordingly, the same security evaluation regarding IND criterion can be employed (including the reduction from an IND-CPA attack to solving the LPN problem) resulting in the statement of IND security.

4.3 Security Implied by Hardness of Recovering Secret Key Based on the Algebraic Representation of Encryption

This section yields a security evaluation via consideration the hardness of recovering the secret key based on the algebraic representation of the encryption in CPA scenario.

Preliminaries

Let's consider encryption of a single plaintext vector $\mathbf{a}$. The following vector equation yields an origin for consideration of hardness of recovering the secret key. For simplicity we assume a chosen plaintext attack where the data are all zeros, i.e. $\mathbf{a} = \mathbf{0}$, and according to (11) we have the following:

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix} \oplus \begin{bmatrix} \mathcal{L}_1(\{r_i\}_i) \\ \mathcal{L}_2(\{r_i\}_i) \\ \cdot \\ \cdot \\ \cdot \\ \mathcal{L}_n(\{r_i\}_i) \end{bmatrix} \oplus \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{bmatrix} \quad , \tag{14}$$

where $\mathcal{L}_i(\cdot)$, $i = 1, 2, ..., n$, are certain linear functions of the arguments $\{r_i\}_i$.

Note that in the set $\{\mathcal{L}_i(\cdot)\}_{i=1}^n$ the elements could be split into two non-overlapping subsets such that one subset contains $m - \ell$ mutually independent elements, and the other subset contains $n - m + \ell$ elements each of which is a linear combination of the elements from the first set. Consequently and assuming for simplicity of a preliminary consideration that the matrix $\mathbf{G}_E$ is an identity matrix, the previous vector equation implies the following one.

$$\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ z_n \end{bmatrix} \oplus \begin{bmatrix} r_1 \\ r_2 \\ . \\ . \\ . \\ r_{m-\ell} \\ \mathcal{L}'_{m-\ell+1}(\{r_i\}_i) \\ \mathcal{L}'_{m-\ell+2}(\{r_i\}_i) \\ . \\ . \\ . \\ \mathcal{L}'_n(\{r_i\}_i) \end{bmatrix} \oplus \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ v_n \end{bmatrix}$$

where $\mathcal{L}'_i(\cdot)$, $i = m-\ell+1, m-\ell+2, ..., n$, are certain linear functions of the arguments $\{r_i\}_i$. Accordingly, note that the vector $[r_1, r_2, \cdots, r_{m-\ell}, \mathcal{L}'_{m-\ell+1}(\{r_i\}_i), \mathcal{L}'_{m-\ell+2}(\{r_i\}_i), \cdots, \mathcal{L}'_n(\{r_i\}_i)]$ contains $m-\ell$ independent elements and another $n-m+\ell$ which are certain linear combinations of the first $m-\ell$ elements.

The previous is a direct consequence of the following: For $n > m-\ell$ there are just $m-\ell$ independent realizations $\{r_i\}_{i=1}^{m-\ell}$ of the random variables R_i, $\Pr(R_i = 1) = \Pr(R_i = 1) = 1/2$, $i = 1, 2, ..., m-\ell$. Accordingly, always exists a vector $[\alpha_1, \alpha_2, ..., \alpha_n] \in \{0,1\}^n$ such that

$$\bigoplus_{i=1}^{n} \alpha_i (\bigoplus_{j=1}^{m-\ell} g_{\ell+j,i} r_j) = 0 \tag{15}$$

The previous implies that for each $i = 1, 2, ..., n$, we have the following equation:

$$x_i \oplus \bigoplus_{j=1,\, j\neq i}^{n} \alpha_j x_j = (z_i \oplus \bigoplus_{j=1,\, j\neq i}^{n} \alpha_j z_j) \oplus (v_i \oplus \bigoplus_{j=1,\, j\neq i}^{n} \alpha_j v_j) \; . \tag{16}$$

Finally note the following: Via recording a sequence of elementary encryption cycles for the time instances $t = 1, 2, ..., \tau$, and assuming that at each time instance the vector $\mathbf{x}^{(t)} = \mathbf{u}^{(t)}\mathbf{S}$, we record the sample for cryptanalysis $\{\mathbf{z}^{(t)}\}_{t=1}^{\tau}$ such that the following is valid:

$$\mathbf{z}^{(t)} = [\mathbf{a}^{(t)} || \mathbf{r}^{(t)}]\mathbf{G} \oplus \mathbf{x}^{(t)} \oplus \mathbf{v}^{(t)} \; , \;\; t = 1, 2, ..., \tau \; . \tag{17}$$

Security Claim
Regarding the following security claim we assume that the probability p of ones in the noise vector $\mathbf{v}$ is adjusted to the error-correction capability of the employed error-correction code.

Theorem 2. When a sample of $n\tau$ ciphertext bits is available in CPA scenario, secret key recovery based on algebraic representation of the proposed encryption scheme is as hard as solving the $\mathrm{LPN}_{kn,q,\epsilon}$ problem with $\epsilon = \frac{1-(1-2p)^{(m-\ell)/2}}{2}$ when $q = (n-m+\ell)\tau$ queries are involved, where p is the crossover probability of a binary symmetric channel for which the employed error-correction code is designed, and ℓ, m, k, and n are the parameters.

The proof of Theorem 2 is given in the Appendix.

Particularly note that involvement of the noise $\{\mathbf{v}^{(t)}\}_t$ prevents employment of the time-memory and time-memory-data trade-off techniques (see [8], [2] and [13], for example) for a more efficient exhaustive search recovery of the secret key.

5 Complexity of Implementation and Communications Overhead

5.1 Complexity of Implementation

Preliminary Note. Following [1], in order to get a quasilinear time implementation of the error correction coding (for sufficiently long messages), the codes reported in [18], with the property that the encoding can be computed via a circuit of size $O(m)$ and the decoding can be decoded by a circuit of size $O(m\log_2 m)$ can be employed.

Complexity Estimation. Complexity of encrypting/decrypting a bit of plaintext (the normalized complexity) can be directly estimated as follows.

- Complexity of Encryption: It is dominantly determined by the sum of the complexities of the employed homophonic encoding, error-correction encoding, and keystream generation.
  - Complexity of homophonic encoding: $O(m)/\ell$ (assuming employment of a sparse/low-density generator matrix)
  - Complexity of error-correction encoding: $O(n)/\ell$ (assuming employment of a sparse/low-density generator matrix)
  - Complexity of keystream generation (the vector-matrix multiplication $\mathbf{uS}$): $(k \cdot n)/\ell$.

  Overall normalized complexity of encryption: $O(m)/\ell + O(n)/\ell + (k \cdot n)/\ell \approx (k \cdot n)/\ell$ assuming that $k > 100$, which is a typical case (see [7], for example).
- Complexity of Decryption: It is dominantly determined by the sum of the complexities of the keystream generation, employed error-correction encoding, and employed homophonic encoding.
  - Complexity of keystream generation (the vector-matrix multiplication $\mathbf{u} \cdot \mathbf{S}$): $(k \cdot n)/\ell$.
  - Complexity of error-correction decoding: $O(n)/\ell$ (assuming employment of a suitable low-complexity decoding code)
  - Complexity of homophonic decoding: $O(m)/\ell$ (assuming employment of a sparse/low-density generator matrix which inverse is a sparse/low-density matrix as well)

  Overall normalized complexity of decryption: $(k \cdot n)/\ell + O(n)/\ell + O(m)/\ell \approx (k \cdot n)/\ell$ when $k > 100$, which is a typical setting.

5.2 Communications Overhead

Communications overhead per-a-bit of plaintext is an implication of: (i) requirement for availability of the pair $(\mathbf{u}, \mathbf{z})$ for the decryption purposes; (ii) employment of homophonic and error-correction coding. Let α_H and α_E denote the communications overhead rates due to the homophonic and error-correction encoding, respectively.

The overall communications overhead α is:

$$\alpha = \frac{k}{\ell} + \alpha_H \cdot \alpha_E = \frac{k}{\ell} + \frac{m}{\ell} \cdot \frac{n}{m} = \frac{k+n}{\ell} \ . \tag{18}$$

6 Comparison with the Related Encryption Schemes

The encryption scheme proposed in this paper originates and follows the paradigm of the ones reported in [7] and [1] and accordingly, all three schemes have the following common feature: (i) they are symmetric encryption based on the LPN problem hardness; (ii) they employ deliberate noise (from a dedicated source of randomness) and error-correction coding; (iii) they are based on simple binary additions/multiplications and vector/matrix operations. Additionally, the scheme proposed in this paper employs homophonic coding in order to provide an enhanced security and additional flexibility of the parameters selection, as well as a possibility for trade-off between the security, communications overhead and implementation complexity.

6.1 Security Issues Related to the Underlying Ideas and the LPN Problems

The schemes reported in [7] and [1] as well as the scheme proposed in this paper are based on the LPN problem which provides that the security appears as the hardness of solving the LPN problem. The approach proposed in this paper provides an enhanced hardness of the underlying LPN problem (even for certain reduced values of the parameters of the LPN problem). The enhancing appears as a consequence of employment a dedicated homophonic encoding which involves pure randomness. Involvement of the homophonic encoding is done by replacement of the error-correction block by a concatenation of the homophonic and error-correction encoding. It has been shown that employment of homophonic encoding implies amplification of the initial noise involved in the LPN problem. Accordingly, the security of the previously reported schemes and the proposed one corresponds to different LPN problems. Let's the $\text{LPN}_{k,n,\epsilon}$ and $\text{LPN}_{k^*,n^*,\epsilon^*}$ denote the LPN problems corresponding to the schemes reported in [7] and [1], and the proposed scheme, respectively. Theorem 2 implies that $\epsilon^* = \frac{1-(1-2p)^{(m-\ell)/2}}{2}$. On the other hand according to the best known algorithms for solving the LPN problem [11] and [4], the hardness of solving the LPN problem heavily depends on the parameter ϵ and when it increase the complexity of solving the LPN problem heavily increase. Accordingly, for the same parameters k, n, the proposed scheme provides substantially higher security level in comparison with the previously reported ones which employ $\epsilon = p$.

The proposed encryption involves balanced random bits (which are not known in advance at the receiver side) in each ciphertext bit: In order to easily learn these balanced random bits, the secret must be known. The homophonic encryption scheme provides involvement of the pure randomness into each bit of the ciphertext which can be easily removed when the secret key is known but removing these balanced random bits from the ciphertext without knowledge of a secret key is as hard as solving certain LPN problem. Finally, note the following: (i) The algebraic representation of proposed encryption shows that each ciphertext bit is affected with $\frac{m-\ell}{2}$ purely random bits which are realizations of a random process which generates zeros and ones independently and with the probability equal to $\frac{1}{2}$; (ii) no one purely random bit is involved into a ciphertext bit in the schemes from [7] and [1].

It has been shown that all three schemes, the ones reported in [7] and [1] as well as the proposed one are IND-CPA secure assuming that the underlying LPN problem is hard one. Note that CCA security of the proposed scheme can be achieved in the same manner as discussed in [7] for the there reported scheme. (The most straightforward way to get an encryption scheme secure against chosen-ciphertext attacks from an encryption scheme secure against chosen-plaintext attacks is to add message authenticity, e.g. by using a Message Authentication Code (MAC)).

As an additional security requirement beside the CPA related indistinguishability one, this paper considers the security of the proposed encryption implied by hardness of recovering the secret key via processing the system of algebraic equations which correspond to the encryption process.

Accordingly, the security goals regarding the proposed scheme are oriented towards CPA security, and the security related to the analysis of complexity of a generic algebraic attack mounted over the algebraic representation of the scheme.

Accordingly, the main security features of the proposed scheme and the corresponding ones from [7] and [1] are compared in Table 1. The illustrative numerical values given in Table 1 corresponds to the results reported in [11] and [7].

Finally note that one of the main roles of the encryption schemes from [1] is to provide security against certain key-dependent message (KDM) attacks; that is, they remain secure even when the adversary is allowed to obtain encryptions of messages that depend on the secret keys themselves, via any affine function of the adversary's choice. Consideration of the KDM security of the proposed encryption is out of the scope of this paper.

Table 1. A comparison of certain features of the proposed encryption and two related ones recently reported in [7] and [1]. (The "balanced random bit" is one which takes values "0" and "1" with the same probability equal to 1/2.)

	parameters of the underlying LPN problem	expected # of unknown balanced random bits involved in a ciphettext bit
symmetric encryptions [7] & [1]	k, n, ϵ	0
proposed encryption	k^*, n^*, $\epsilon^* = \frac{1-(1-2p)^{(m-\ell)/2}}{2}$ typically: $k^* << k$, $n^* \approx n$, $p << \epsilon$	$(m-\ell)/2$

6.2 Comparison of the Communications Overhead and Implementation Complexity

The normalized communications overheads of the cipher [7] (and similarly for the corresponding cipher [1]) and the proposed one are $(k+n)/\ell$ and $(k^*+n^*)/\ell$, respectively, noting that typically $k^* << k$ and $n^* \approx n$ because a higher rate error-correcting code could be employed because the same security level could be obtained with a smaller p which provides $\epsilon^* > \epsilon$. This higher rate error-correcting code can compensate the overhead implied by the employed homophonic coding.

Regarding the implementation complexity note that the proposed scheme requires only one additional (homophonic) encoding at the encryption side and one additional (homophonic) decoding at the decryption side. When the employed homophonic code is a linear one, the encoding and decoding are vector-matrix multiplications. As discussed in Section 5.1, if a linear time encoding/decoding error-correction code is employed (as suggested in [1]) the implementation complexities of the proposed scheme could be summarized as follows: (i) Overall normalized complexity of encryption: $\approx (k^* \cdot n^*)/\ell$ assuming that $k* > 100$, which is a typical case (see [7], for example); (ii) Overall normalized complexity of decryption: $\approx (k^* \cdot n^*)/\ell$ when $k^* > 100$.

The implementation complexity of the scheme [7] can be estimated as follows: (i) Normalized complexity of generating a ciphertext is dominated by the error-correction encoding and encryption and so it is $O(n)/\ell + (k \cdot n)/\ell \approx (k \cdot n)/\ell$ assuming that $k > 100$, which is a typical case (see [7], for example); (ii) Normalized complexity of recovering the plaintext from the ciphertext dominated by the decryption and error-correction decoding and so it is $(k \cdot n)/\ell + O(n)/\ell \approx (k \cdot n)/\ell$ mod2 additions when $k > 100$. A similar estimation of the normalized complexities can be done for the considered symmetric encryption [1].

The considered LPN-based constructions mainly rely on addition and multiplication of large binary vectors / matrices. These operations can be performed very fast in practice even if one does not employ the asymptotically-fast algorithms. It is claimed in [1] that in particular, as in the case of the HB protocol (see [9] and [6], for example), that the considered LPN problem based schemes (or variants of them) might turn to be useful for hardware implementation by computationally-weak devices.

Table 2 yields an illustrative numerical comparison of certain issues relevant for consideration of the implementation complexity.[2]

[2] In details discussion of the error-correcting coding issues is out of the scope of this paper Particularly, just note that the choice of the noise parameter p and the code influences not only security, but also the probability that decryption is successful: Accordingly the comparisons of different settings should assume similar decryption success.

Table 2. A comparison of a normalized implementation complexity (complexity of a bit encryption/decryption) and the communications overhead of the proposed encryption and the one reported in [7] (and similarly for the symmetric one recently reported in [1].) noting that typically $k^* << k$ and $n^* \approx n$.

	normalized implementation complexity	communications overhead	illustrative numerical values of the parameters
encryption [7]	$\sim kn/\ell$	$(k+n)/\ell$	$\epsilon = p = 0.05$ $\ell = 75$, $k = 768$, $n = 160$
proposed encryption	$\sim k^*n^*/\ell$	$(k^*+n^*)/\ell$	$m-\ell = 30$, $p = 0.025$, $\epsilon^* = 0.268$ $\ell = 75$, k^*512, $n^* = 160$

7 Concluding Summary

A novel approach for design of certain ciphers has been proposed which is based on joint employment of pseudorandomness, randomness and dedicated coding. The pseudorandomness is generated by a known vector and secret matrix multiplication over GF(2). Two types of binary randomness are employed: The pure one where independent bits have the same probabilities of zeros and ones, and a biased one where the independent bits take the value 1 with the probability $p << 1/2$ and the value 0 with the probability $1-p$. Also two types of dedicated coding are employed: One dedicated coding belongs to the class of the homophonic (i.e. wire-tap channel) coding techniques, and the other one is the error-correcting code. The employed homophonic encoding provides involvement of the pure randomness in the ciphertext and it is such that provides low-complexity extraction of the randomness, i.e. decoding, when the secret key is known, and at he same time makes the decoding as complex as exhaustive search over all possible secret keys when the secret key is not known. Accordingly, the homophonic encoding provides a heavy uncertainty at the side of an attacker. The employed error-correction coding is such that provides error-free decoding after a binary symmetric channel with the crossover probability p. Algebraic representation of the proposed cipher has been employed for the security evaluation and analysis of the implementation complexity and the communications overhead.

The security evaluation of the proposed encryption is considered from the following two points of view: (i) security implied by the indistinguishability (IND) in the chosen plaintext attack (CPA) scenario assuming hardness of the underlying LPN problem; (ii) security implied by hardness of recovering the secret key based on the algebraic representation of the encryption in CPA scenario. Regarding (i), it is shown that the encryption is IND CPA secure, and regarding (ii) it is shown that the addressed secret key recovery is as hard as the LPN when the corrupting noise is $\epsilon = \frac{1-(1-2p)^{(m-\ell)/2}}{2}$ and m, ℓ and p are the stream cipher parameters. Accordingly, assuming that the parameters of the scheme are appropriately selected, the complexity of the secret key recovery based on the algebraic representation is approximately as hard as the exhaustive search over all possible secret keys. Particularly note that IND-CPA security appears as an implication of the assumption on hardness of the underlying LPN problem, and it should be taken into account that in certain settings the LPN problem is much easier than in the worst case scenario, and accordingly additional security considerations, like the above (ii) one, are desirable. Finally note that the homophonic encoding transforms an LPN type problem with a certain noise rate into another LPN type problem with a larger noise rate, thereby making the problem harder.

Consideration of the implementation complexity shows that the implementation complexity is low regarding the both time and space, assuming that the linear block codes are employed. In a number of settings the dominant operations are vector/matrix multiplications over GF(2).

Finally, the proposed cipher design is compared with the related symmetric encryptions reported in [7] and [1] which are origins of the proposed scheme. The comparison implies that the implementation complexities and the communications overheads in all three considered encryption schemes are similar when the same parameters are employed. On the other hand, the proposed scheme can provide the same level of security for reduced values of certain parameters implying that, in certain scenarios, it can provide lower communications overhead and more efficient implementation. Also, all three schemes fulfil the indistinguishability security criterion in CPA. On the other hand, the proposed scheme provide the enhanced security regarding the complexity of recovering the secret key via processing the algebraic equations which represent the encryption process. For all three schemes this recovery is as hard as solving the LPN_ϵ problem with $\epsilon = p$ for the schemes [7] and [1], and $\epsilon = \frac{1-(1-2p)^{(m-\ell)/2}}{2}$ which implies an enhanced security regarding the algebraic key recovery. The homophonic encryption scheme provides involvement of the pure randomness into each bit of the ciphertext which can be easily removed when the secret key is known but removing these balanced random bits from the ciphertext without knowledge of a secret key is as hard as solving certain LPN problem.

8 Appendix: Proof of Theorem 2

In the considered CPA scenario which corresponds to the all zeros plaintext, for each t, $1 \leq t \leq \tau$, a single ciphetext word implies the following system of a basic equations when the plaintext consists of all zeros:

$$\begin{array}{rcl}
x_1^{(t)} & = & z_1^{(t)} \oplus \mathcal{L}_1(\{r_i^{(t)}\}_i) \oplus v_1^{(t)} \\
x_2^{(t)} & = & z_2^{(t)} \oplus \mathcal{L}_2(\{r_i^{(t)}\}_i) \oplus v_2^{(t)} \\
\cdot & & \\
\cdot & & \\
\cdot & & \\
x_{m-\ell}^{(t)} & = & z_{m-\ell}^{(t)} \oplus \mathcal{L}_{m-\ell}(\{r_i^{(t)}\}_i) \oplus v_{m-\ell}^{(t)} \\
\hline
x_{m-\ell+1}^{(t)} & = & z_{m-\ell+1}^{(t)} \oplus \mathcal{L}_{m-\ell+1}(\{r_i^{(t)}\}_i) \oplus v_{m-\ell+1}^{(t)} \\
x_{m-\ell+2}^{(t)} & = & z_{m-\ell+2}^{(t)} \oplus \mathcal{L}_{m-\ell+2}(\{r_i^{(t)}\}_i) \oplus v_{m-\ell+2}^{(t)} \\
\cdot & & \\
\cdot & & \\
\cdot & & \\
x_n^{(t)} & = & z_n^{(t)} \oplus \mathcal{L}_n(\{r_i^{(t)}\}_i) \oplus v_n^{(t)}
\end{array} \tag{19}$$

where $\mathcal{L}_i(\cdot)$, $i = 1, 2, ..., n$, are certain linear operators.

Via suitable linear combining of the above equations we obtain the following processed system of equations related to a single word when the plaintext consists of all zeros.

$$
\begin{array}{ccccccc}
x_1^{(t)} & = & z_1^{(t)} & \oplus & \mathcal{L}_1(\{r_i^{(t)}\}_i) & \oplus & v_1^{(t)} \\
x_2^{(t)} & = & z_2^{(t)} & \oplus & \mathcal{L}_2(\{r_i^{(t)}\}_i) & \oplus & v_2^{(t)} \\
\cdot & & & & & & \\
\cdot & & & & & & \\
\cdot & & & & & & \\
x_{m-\ell}^{(t)} & = & z_{m-\ell}^{(t)} & \oplus & \mathcal{L}_{m-\ell}(\{r_i^{(t)}\}_i) & \oplus & v_{m-\ell}^{(t)} \\
\hline
\mathcal{L}^*_{m-\ell+1}(\{x_i^{(t)}\}_i) & = & \mathcal{L}^*_{m-\ell+1}(\{z_i^{(t)}\}_i) & \oplus & 0 & \oplus & \mathcal{L}^*_{m-\ell+1}(\{v_i^{(t)}\}_i) \\
\mathcal{L}^*_{m-\ell+2}(\{x_i^{(t)}\}_i) & = & \mathcal{L}^*_{m-\ell+2}(\{z_i^{(t)}\}_i) & \oplus & 0 & \oplus & \mathcal{L}^*_{m-\ell+2}(\{v_i^{(t)}\}_i) \\
\cdot & & & & & & \\
\cdot & & & & & & \\
\cdot & & & & & & \\
\mathcal{L}^*_n(\{x_i^{(t)}\}_i) & = & \mathcal{L}^*_n(\{z_i^{(t)}\}_i) & \oplus & 0 & \oplus & \mathcal{L}^*_n(\{v_i^{(t)}\}_i)
\end{array}
\tag{20}
$$

where $\mathcal{L}^*_i(\cdot)$, $i = 1, 2, ..., n$, are certain linear operators.

Accordingly, discarding the equations which contain the random bits, i.e. the components $\mathcal{L}_i(\{r_j^{(t)}\}_j)$, $i = 1, 2, ..., m-\ell$, we obtain the following aggregated system of $(n-m+\ell)\tau$ equations corresponding to $t = 1, 2, ..., \tau$.

$$
\begin{array}{ccccc}
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
\mathcal{L}^*_{m-\ell+1}(\{x_i^{(t)}\}_i) & = & \mathcal{L}^*_{m-\ell+1}(\{z_i^{(t)}\}_i) & \oplus & \mathcal{L}^*_{m-\ell+1}(\{v_i^{(t)}\}_i) \\
\mathcal{L}^*_{m-\ell+2}(\{x_i^{(t)}\}_i) & = & \mathcal{L}^*_{m-\ell+2}(\{z_i^{(t)}\}_i) & \oplus & \mathcal{L}^*_{m-\ell+2}(\{v_i^{(t)}\}_i) \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
\mathcal{L}^*_n(\{x_i^{(t)}\}_i) & = & \mathcal{L}^*_n(\{z_i^{(t)}\}_i) & \oplus & \mathcal{L}^*_n(\{v_i^{(t)}\}_i) \\
\hline
\mathcal{L}^*_{m-\ell+1}(\{x_i^{(t+1)}\}_i) & = & \mathcal{L}^*_{m-\ell+1}(\{z_i^{(t+1)}\}_i) & \oplus & \mathcal{L}^*_{m-\ell+1}(\{v_i^{(t+1)}\}_i) \\
\mathcal{L}^*_{m-\ell+2}(\{x_i^{(t+1)}\}_i) & = & \mathcal{L}^*_{m-\ell+2}(\{z_i^{(t+1)}\}_i) & \oplus & \mathcal{L}^*_{m-\ell+2}(\{v_i^{(t+1)}\}_i) \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
\mathcal{L}^*_n(\{x_i^{(t+1)}\}_i) & = & \mathcal{L}^*_n(\{z_i^{(t+1)}\}_i) & \oplus & \mathcal{L}^*_n(\{v_i^{(t+1)}\}_i) \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & &
\end{array}
\tag{21}
$$

According to (11) we have

$$
x_i^{(t)} = (\bigoplus_{j=1}^{k} s_{j,i} u_j^{(t)}) \ , \quad i = 1, 2, ..., n \tag{22}
$$

and accordingly, the system of equations (21) becomes the following one:

$$
\begin{array}{rcl}
\vdots & & \\
\mathcal{L}^*_{m-\ell+1}(\{\bigoplus_{j=1}^k s_{j,i}u_j^{(t)}\}_i) & = & \mathcal{L}^*_{m-\ell+1}(\{z_i^{(t)}\}_i) \oplus \mathcal{L}^*_{m-\ell+1}(\{v_i^{(t)}\}_i) \\
\mathcal{L}^*_{m-\ell+2}(\{\bigoplus_{j=1}^k s_{j,i}u_j^{(t)}\}_i) & = & \mathcal{L}^*_{m-\ell+2}(\{z_i^{(t)}\}_i) \oplus \mathcal{L}^*_{m-\ell+2}(\{v_i^{(t)}\}_i) \\
\vdots & & \\
\mathcal{L}^*_n(\{\bigoplus_{j=1}^k s_{j,i}u_j^{(t)}\}_i) & = & \mathcal{L}^*_n(\{z_i^{(t)}\}_i) \oplus \mathcal{L}^*_n(\{v_i^{(t)}\}_i) \\
\hline
\mathcal{L}^*_{m-\ell+1}(\{\bigoplus_{j=1}^k s_{j,i}u_j^{(t+1)}\}_i) & = & \mathcal{L}^*_{m-\ell+1}(\{z_i^{(t+1)}\}_i) \oplus \mathcal{L}^*_{m-\ell+1}(\{v_i^{(t+1)}\}_i) \\
\mathcal{L}^*_{m-\ell+2}(\{\bigoplus_{j=1}^k s_{j,i}u_j^{(t+1)}\}_i & = & \mathcal{L}^*_{m-\ell+2}(\{z_i^{(t+1)}\}_i) \oplus \mathcal{L}^*_{m-\ell+2}(\{v_i^{(t+1)}\}_i) \\
\vdots & & \\
\mathcal{L}^*_n(\{\bigoplus_{j=1}^k s_{j,i}u_j^{(t+1)}\}_i) & = & \mathcal{L}^*_n(\{z_i^{(t+1)}\}_i) \oplus \mathcal{L}^*_n(\{v_i^{(t+1)}\}_i) \\
\vdots & &
\end{array}
\tag{23}
$$

When each of $m-\ell$ variables $v_i^{(t)}$ is a realization of a random binary variable $V_i^{(t)}$, such that $\Pr(V_i^{(t)}=1) = 1 - \Pr(V_i^{(t)}=0) = p$, $i \in [1,2,...,n]$, $t = 1,2,...$, and assuming that the design of the matrix $\mathbf{G}$ is such that each operator $\mathcal{L}_j^*(\cdot)$ involve a half of them, we have the following (see [5], for example):

$$
\Pr(\mathcal{L}_j^*(\{V_i\}_i) = 1) = \frac{1-(1-2p)^{(m-\ell)/2}}{2} \, . \tag{24}
$$

The system of equations (23) is an overdefined consistent system of $(m-n+\ell)\tau$ linear equations with kn unknown variables where each equations is true with the probability $1 - \frac{1-(1-2p)^{(m-\ell)/2}}{2} = 1-\epsilon$. Note that on the left hand side of each equation is a linear combination of certain unknown and known variables and accordingly it can be considered as a component equation of the $\mathrm{LPN}_{kn,q,\epsilon}$ problem when $q = (m-n+\ell)\tau$ queries are involved.

The above considerations imply the theorem statement.

References

1. B. Applebaum, D. Cash, C. Peikert and A. Sahai, "Fast Cryptographic Primitives and Circular-Secure Encryption Based on Hard Learning Problems", CRYPTO 2009, *Lecture Notes in Computer Science*, vol. 5677, pp. 595-618, Aug. 2009.
2. A. Biryukov and A. Shamir, "Cryptanalytic time/memory/data tradeoffs for stream ciphers", ASIACRYPT 2000, *Lecture Notes in Computer Science*, vol. 1976, pp. 1-13, 2000.
3. A. Blum, M. Furst, M. Kearns and R. Lipton, "Cryptographic Primitives Based on Hard Learning Problems", CRYPTO 1993, *Lecture Notes in Computer Science*, vol. 773, pp. 278291, 1994.
4. M. Fossorier, M.J. Mihaljević, H. Imai, Y. Cui and K. Matsuura, "An Algorithm for Solving the LPN Problem and its Application to Security Evaluation of the HB Protocols for RFID Authentication", INDOCRYPT 2006, *Lecture Notes in Computer Science*, vol. 4329, pp. 48-62, Dec. 2006.
5. M. Fossorier, M.J. Mihaljević and H. Imai, "Modeling Block Encoding Approaches for Fast Correlation Attack", *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4728-4737, Dec. 2007.
6. H. Gilbert, M.J.B. Robshaw and Y. Seurin, "HB$^{\#}$: Increasing the Security and Efficiency of HB^{+}", EUROCRYPT2008, *Lecture Notes in Computer Science*, vol. 4965, pp. 361-378, 2008.
7. H. Gilbert, M.J.B. Robshaw, and Y. Seurin, "How to Encrypt with the LPN Problem", ICALP 2008, Part II, *Lecture Notes in Computer Science*, vol. 5126, pp. 679-690, 2008.

8. M.E. Hellman, “A cryptanalytic time-memory trade-off”, *IEEE Transactions on Information Theory*, vol. 26, pp. 401-406, July 1980.
9. N. Hopper and M. Blum, “Secure Human Identification Protocols”, ASIACRYPT 2001, *Lecture Notes in Computer Science*, vol. 2248, pp. 52-66, 2001.
10. H.N. Jendal, Y.J.B. Kuhn, and J.L. Massey, “An information-theoretic treatment of homophonic substitution”, EUROCRYPT’89, *Lecture Notes in Computer Science*, vol. 434, pp. 382-394, 1990.
11. E. Levieil and P.-A. Fouque, ”An Improved LPN Algorithm”, SCN 2006, *Lecture Notes in Computer Science*, vol. 4116, pp. 348-359, 2006.
12. J. Massey, “Some Applications of Source Coding in Cryptography”, *European Transactions on Telecommunications*, vol. 5, pp. 421-429, July-August 1994.
13. M.J. Mihaljević, M. Fossorier and H. Imai, “Security Evaluation of Certain Broadcast Encryption Schemes Employing a Generalized Time-Memory-Data Trade-Off”, *IEEE Communications Letters*, vol. 11, no. 12, pp. 988-990, Dec. 2007.
14. M.J. Mihaljević and H. Imai, “An approach for stream ciphers design based on joint computing over random and secret data”, *Computing*, vol. 85, no. 1-2, pp. 153-168, June 2009. (DOI: 10.1007/s00607-009-0035-x)
15. M.J. Mihaljević, ”A Framework for Stream Ciphers Based on Pseudorandomness, Randomness and Error-Correcting Coding”, in *Enhancing Cryptographic Primitives with Techniques from Error Correcting Codes*, B. Preneel, at al Eds., Vol. 23 in the Series Information and Communication Security, pp. 117-139, IOS Press, Amsterdam, The Netherlands, June 2009. DOI: 10.3233/978-1-60750-002-5-117 (ISSN: 1874-6268; ISBN: 978-1-60750-002-5)
16. R. Rivest and T. Sherman, ”Randomized Encryption Techniques”, *Advances in Cryptology: Proceedings of CRYPTO ’82*, Plemum, New Yourk, pp. 145-163, 1983.
17. B. Ryabko and A. Fionov, “Efficient Homophonic Coding”, *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 2083-2094, Sept. 1999.
18. D.A. Spielman, ”Linear-time encodable and decodable error-correcting codes”. *IEEE Trans. Information Theory*, vol. 42, No 6, pp. 1723-1732, 1996. (preliminary published in: Proc. 27th STOC, pp. 388-397, 1995)
19. A.D. Wyner, “The wire-tap channel”, *Bell Systems Technical Journal*, vol. 54, pp. 1355-1387, Oct. 1975.