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

Miodrag Mihaljevic and Hideki Imai

Research Center for Information Security (RCIS), National Institute AIST, Tokyo

Symmetric Key Encryption Workshop 2011
Copenhagen, 17 February 2011

Abstract

- This talk 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.
- 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**.

Roadmap

- Introduction
- Encryption Involving Homophonic Coding
- Security Evaluation
- Comparisons
- A Step Forward
- Concluding Remarks

I. Introduction

Encryption Schemes Based on the LPN Problem

Encryption Schemes Based on the LPN Problem

- 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.
- 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.

LPN Problem Based Encryption

To encrypt an ℓ-bit vector $\mathbf{a}$, the sender draws a k-bit random vector $\mathbf{u}$ and computes

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

where $\mathbf{v} \leftarrow \mathrm{Ber}_{n,p}$ is an n-bit noise vector such that each of its bits is (independently) 1 with probability p and 0 with probability $1 - p$.

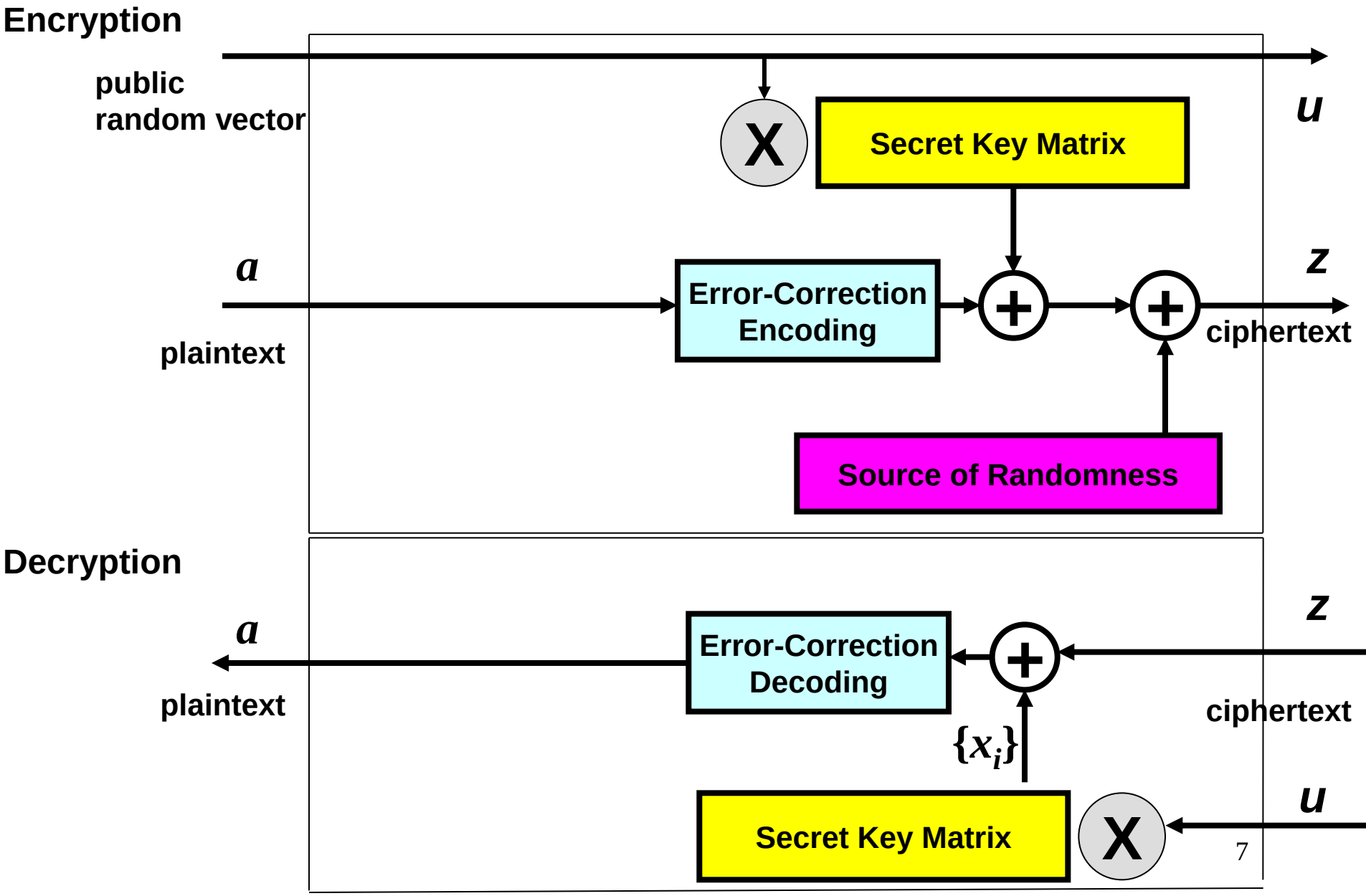
The resulting "ciphertext" is the pair $(\mathbf{u}, \mathbf{z})$.

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.

LPN Problem Based Encryption



II. Encryption Based on Pseudo-Randomness, Randomness and Dedicated Coding

Power of Randomness for Enhancing Security and Low Implementation Complexity

Design Motivations

- 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 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 at ICALP2008 and CRYPTO2009.

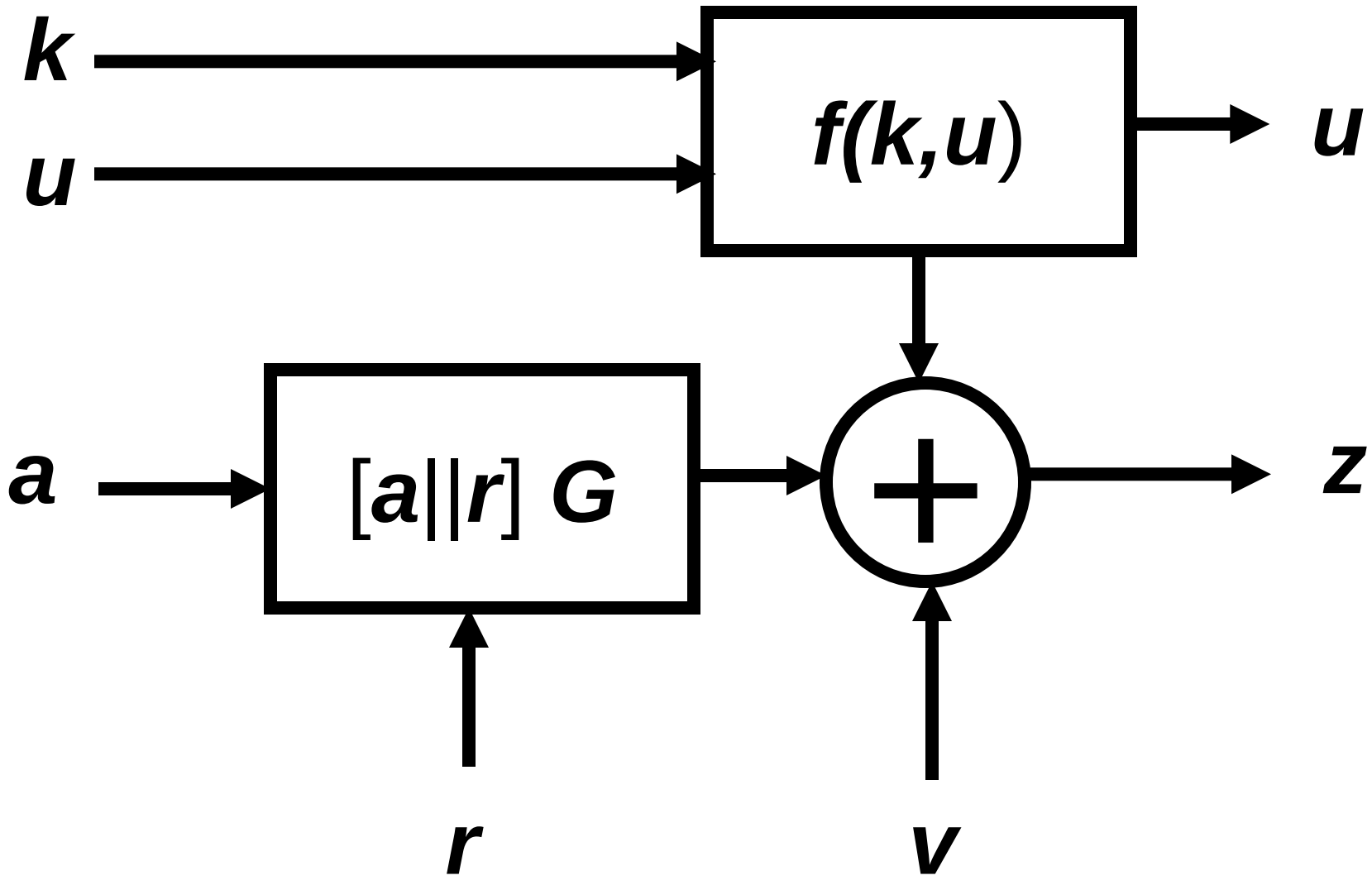
Power of Randomness for High Security and Low Implementation Complexity

Design Components:

- Simple Finite State Machine for the Pseudo-Randomness
- Dedicated Coding: Homophonic and Error-Correction Ones
- Randomness

Effects:

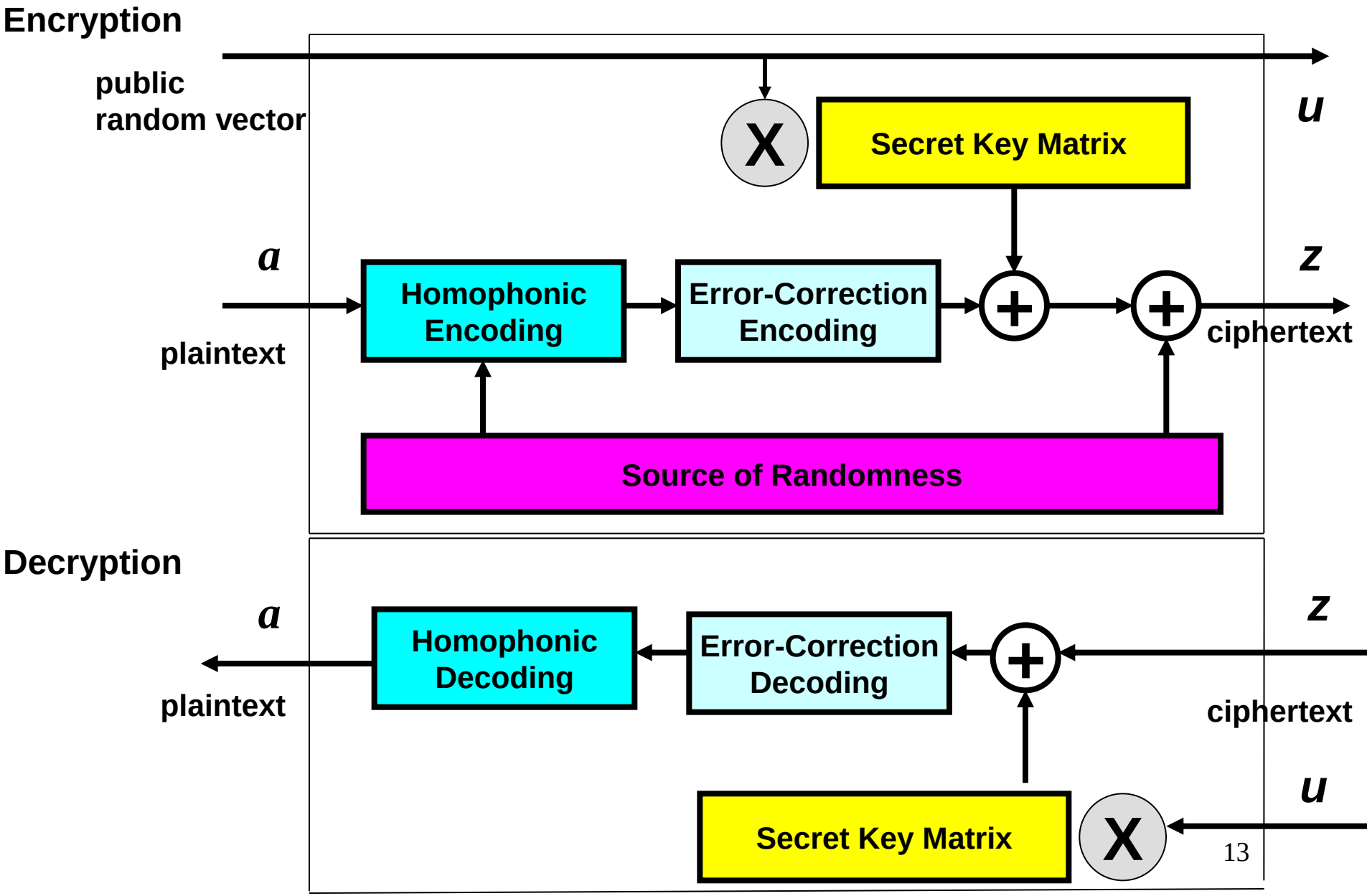
- Enhanced Security Implied by Randomness
- Low Implementation Complexity


k
u
f(k,u)
u
a
[a||r] G
z
r
v

An Advanced LPN Problem Based Encryption Scheme Employing Homophonic Coding

Power of Randomness for Enhancing Security

Homophonic Coding Based LPN Encryption



- *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:

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

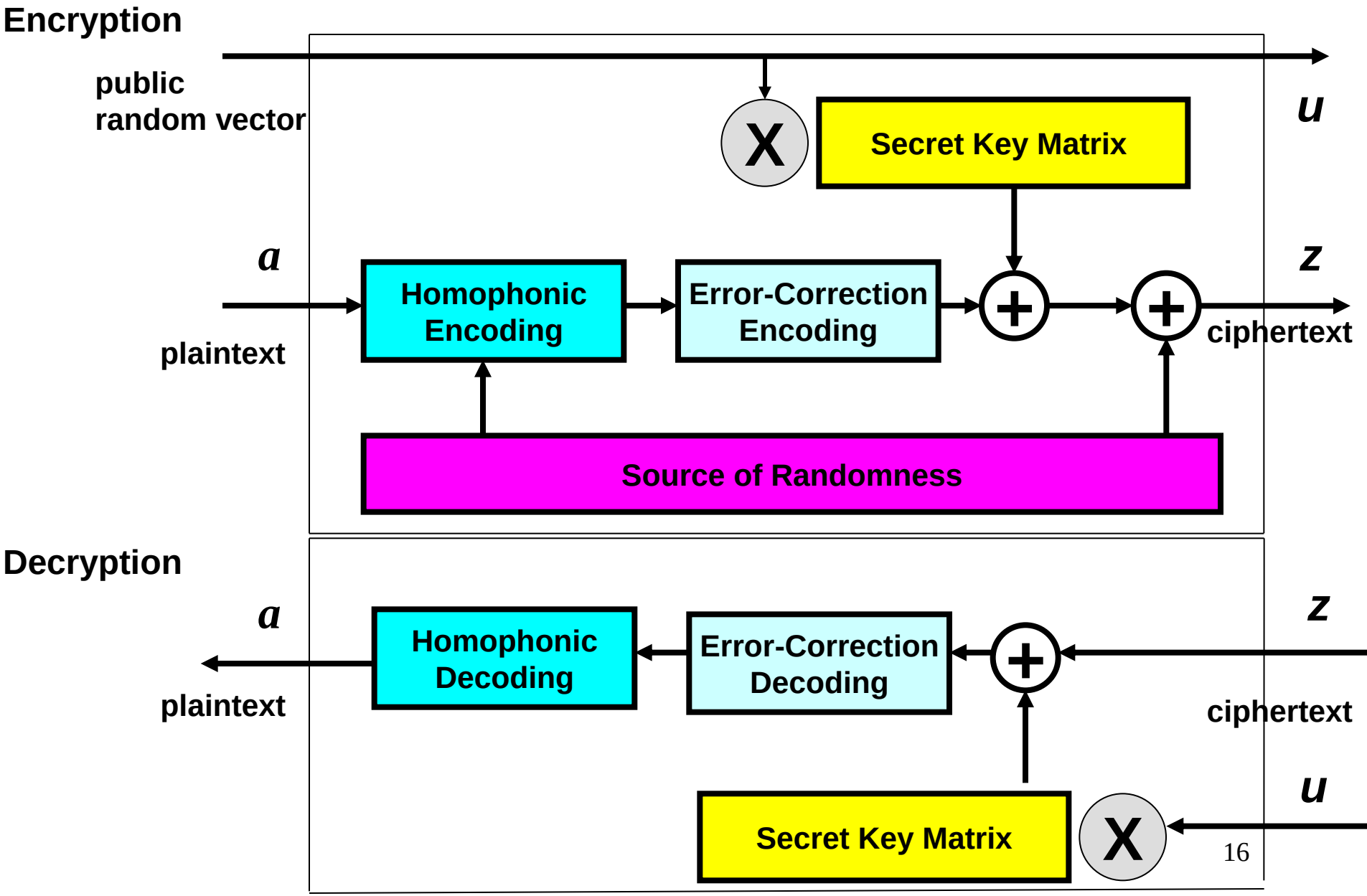
- *Decryption*

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

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

  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.

Homophonic Coding Based LPN Encryption



Homophonic Encoding

Groups of the codewords: *Same symbol denote different codewords belonging to the same group*

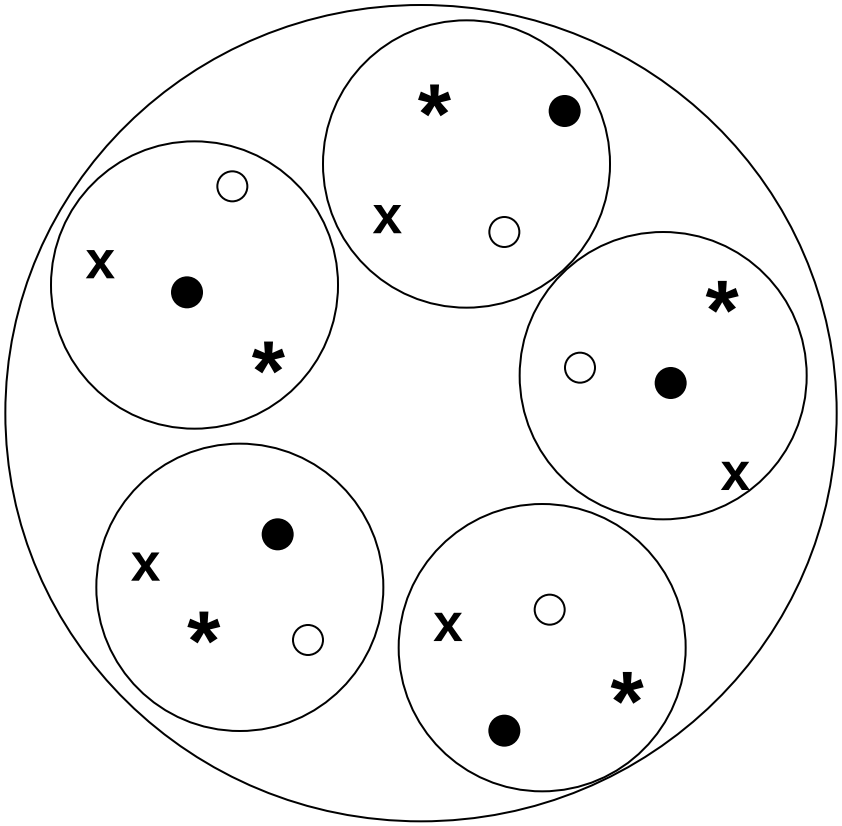
● ● ● ● ●

○ ○ ○ ○ ○

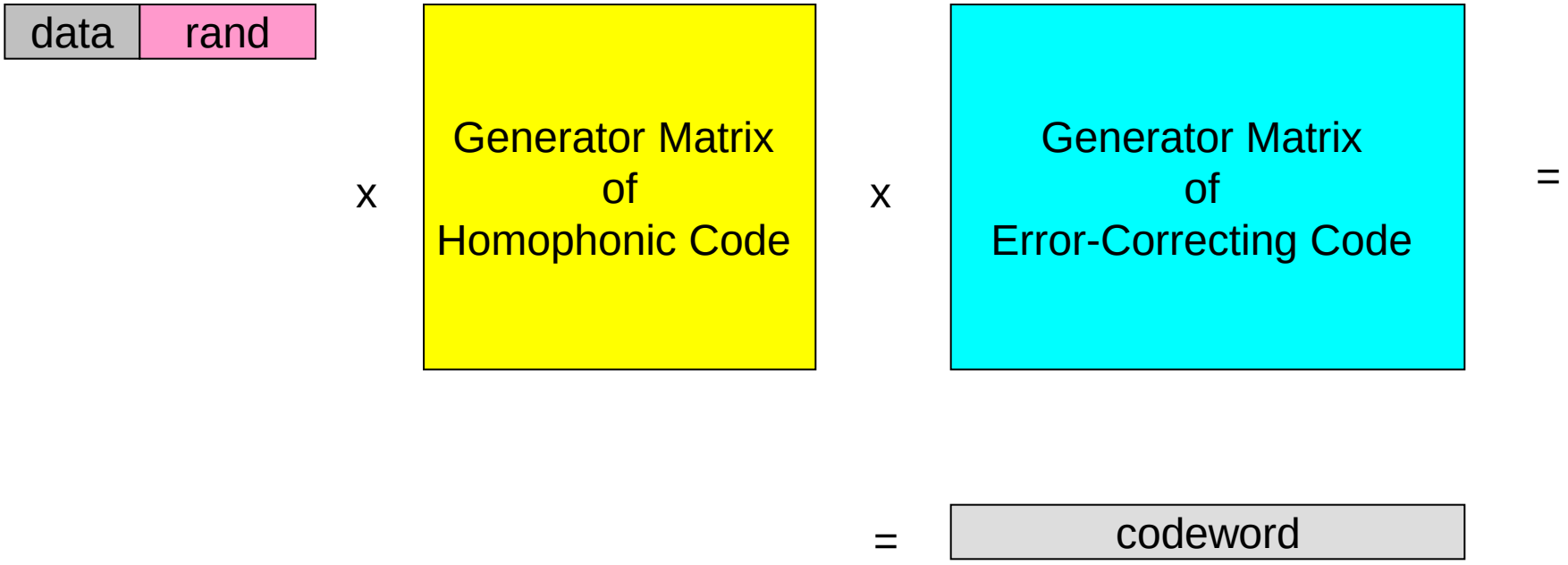
x x x x x

* * * * *

Codewords and N-dim Sphere



Homophonic and Error-Correction Encoding



Origins of for the Enhanced Security

- **Effects of involvement randomness**.
- **Hardness of decoding without secret key.**

III. Security Evaluation

Computational Complexity

Algebraic Representation at

Bit-Level

Let $\mathbf{G} = [g_{i,j}]_{i=1}^{m}{}_{j=1}^{n}$, and let $\mathbf{z} = [z_i]_{i=1}^{n}$. Then,

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

implying that under the known plaintext attack we have

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

where the right-hand side of the equation has known value, and where

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

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

- The Computational Complexity -

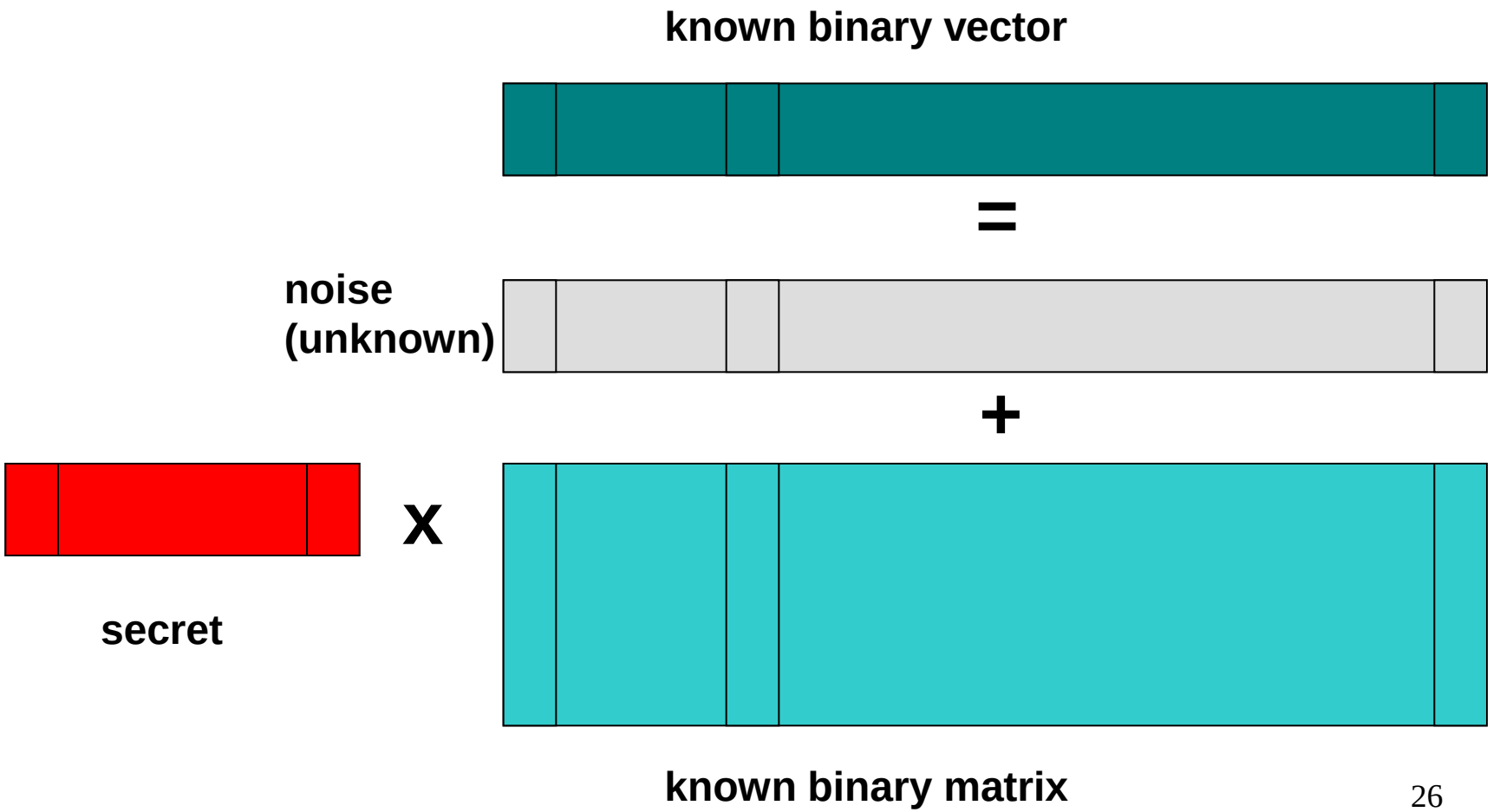
Basic System of Equations Related to a **Single Word** when the Plaintext Consists of all Zeros

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

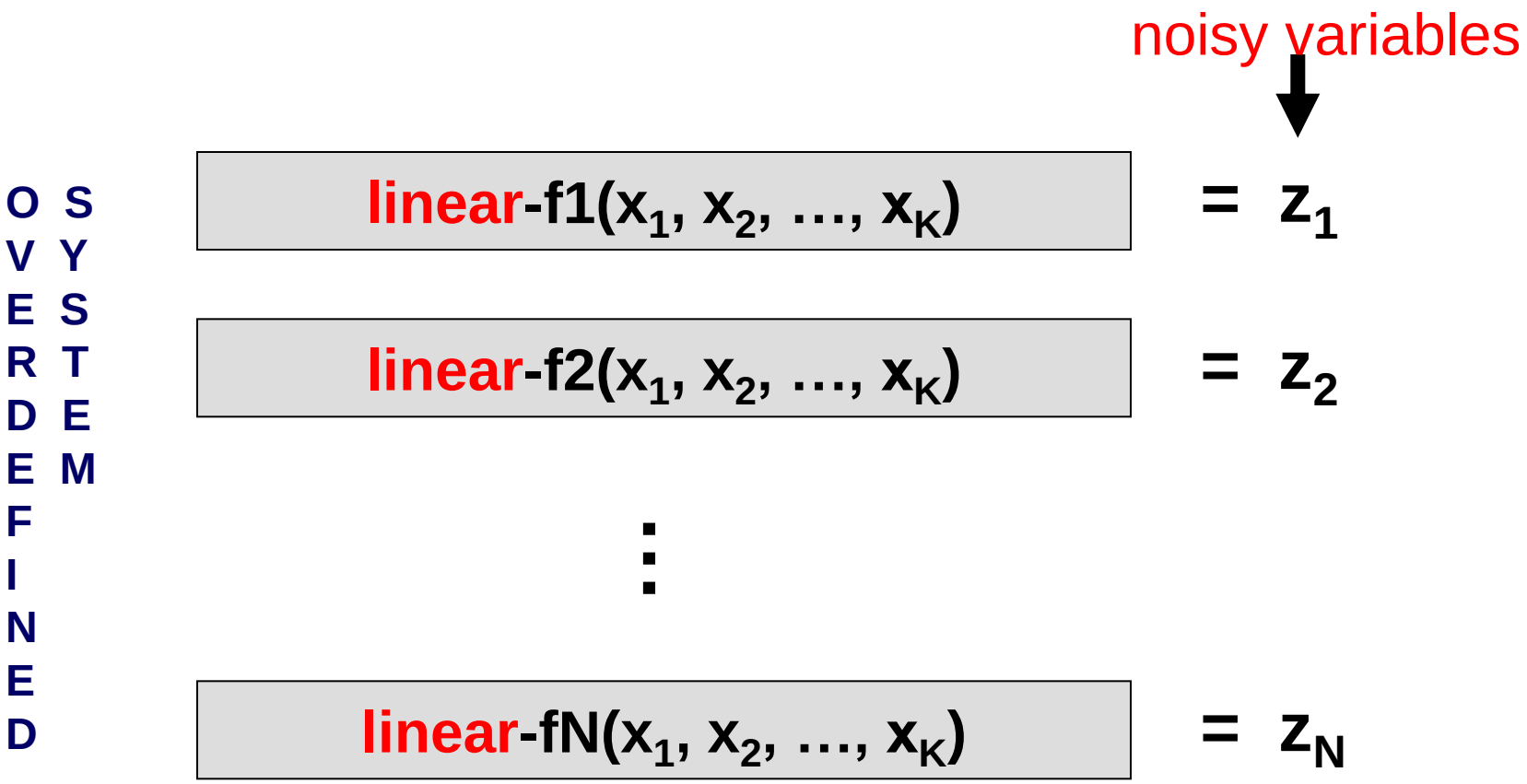
The **Aggregated System** with eliminated "purely random bits"

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

LPN Problem
(an equivalent formulation)



Underlying Problem of the LPN



K << N

The Corrupting Noise

The system of equations implies the following. Assuming that each $v_i^{(t)}$ is a realization of a random binary variable $V_i^{(t)}$, such that $\Pr(V_i^{(t)} = 1) = 1 - \Pr(V_i^{(t)} = 0) = p$ $i = 1, 2, ..., n$, $t = 1, 2, ...$, we have the following:

$$\Pr(\mathcal{L}_j^{(t)}(\{V_i\}_i) = 1) = \frac{1 - (1 - 2p)^w}{2},$$

wherer $w >> 1$ is a parameter.

Security and LPN Problem

Accordingly, the considered system of equations implies the following:

- The security of the considered encryption corresponds to hardness of the LPN problem;
- In the average case scenario the security corresponds to solving the LPN problem (i.e. decoding) when the involved noise has the probability of ones equal to $\frac{1-(1-2p)^w}{2}$ (where w is a parameter implied by the employed homophonic code).

A Claim on Security of the Proposed Encryption

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.

IV. Comparison with the Schemes Reported at

ICALP2008 and CRYPTO2009

A comparison of certain features of the proposed encryption and two related ones recently reported at ICALP2008 and CRYPTO2009. (The "balanced random bit" is one which takes values "0" and "1" withthe 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 ICALP2008 & CRYPTO2009	$k,\ n,\ \epsilon$	0
proposed encryption	$k^*,\ n^*,\ \epsilon^* = \frac{1-(1-2p)^{(m-\ell)/2}}{2}$ typically: $k^* << k,\ n^* \approx n,\ p << \epsilon$	$(m-\ell)/2$

Comparison of Certain Implementation Features

	normalized implementation complexity	communications overhead	illustrative numerical values of the parameters
encryption ICALP2008	$\sim kn/\ell$	$(k+n)/\ell$	$\epsilon = p = 0.05$ $\ell = 75,\ k = 768,\ n = 160$
proposed encryption	$\sim k^*n^*/\ell$	$(k^*+n^*)/\ell$	$m-\ell = 30,\ p = 0.025,\ \epsilon^* = 0.268$ $\ell = 75,\ k^*512,\ n^* = 160$

V. A Step Forward

Homophonic Coding Based Compact Stream Ciphers

Randomized Stream Ciphers

Only Noisy Sample Available for Cryptanalysis

Stream Cipher Approaches

- **One-Time Pad** – pure random approach (provable security)
- Traditional **Keystream Generator** – finite state machine: a deterministic approach (heuristic security)

Randomized approach:

- A stream cipher based on employment of **Pseudorandomness, Randomness and Dedicated Coding**
- **Towards provable security implied by the dimension of secret key**

Power of Randomness for High Security and Low Implementation Complexity

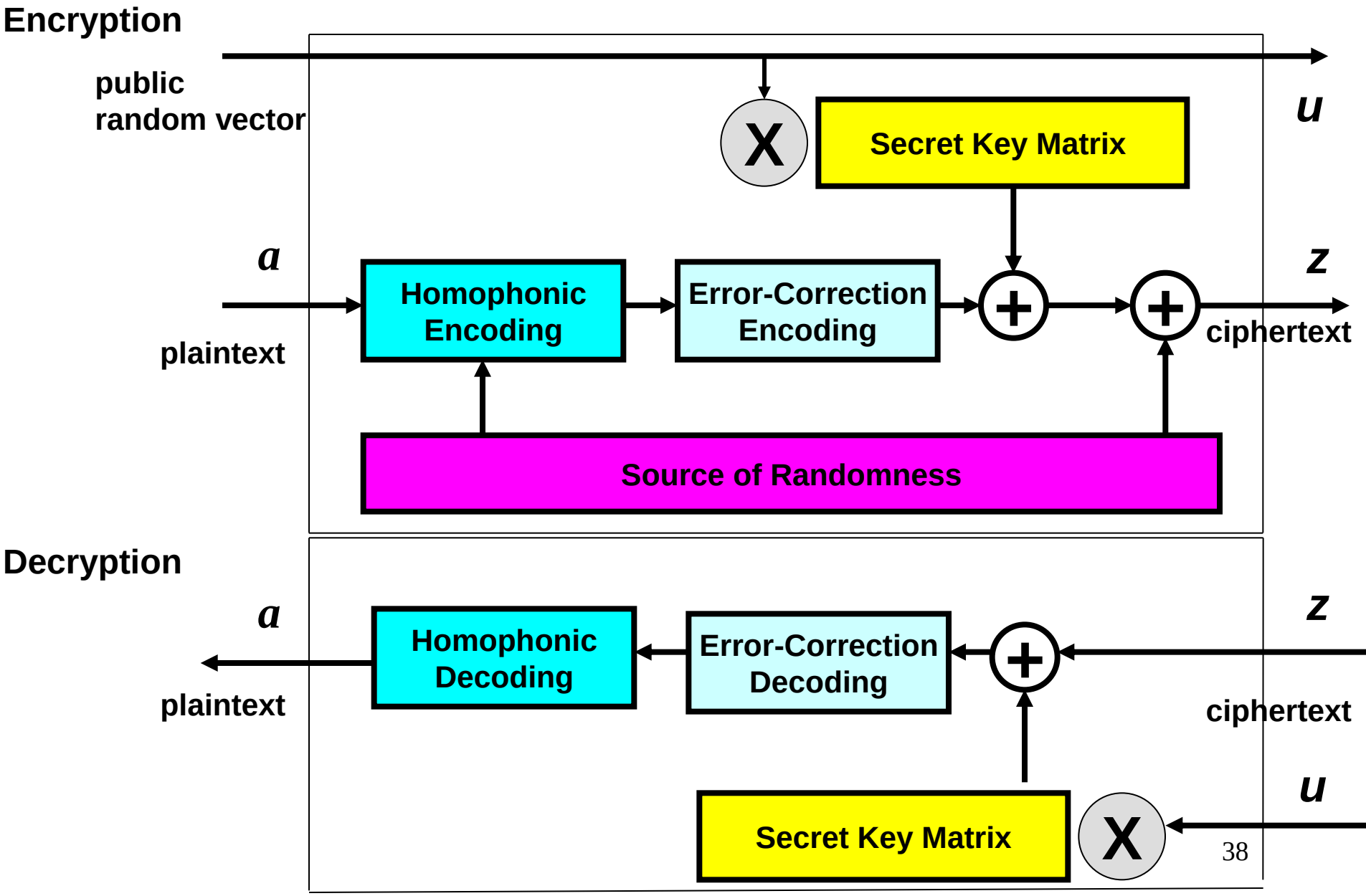
Design Components:

- Simple Finite State Machine for the Pseudo-Randomness
- Dedicated Coding: Homophonic and Error-Correction Ones
- Randomness

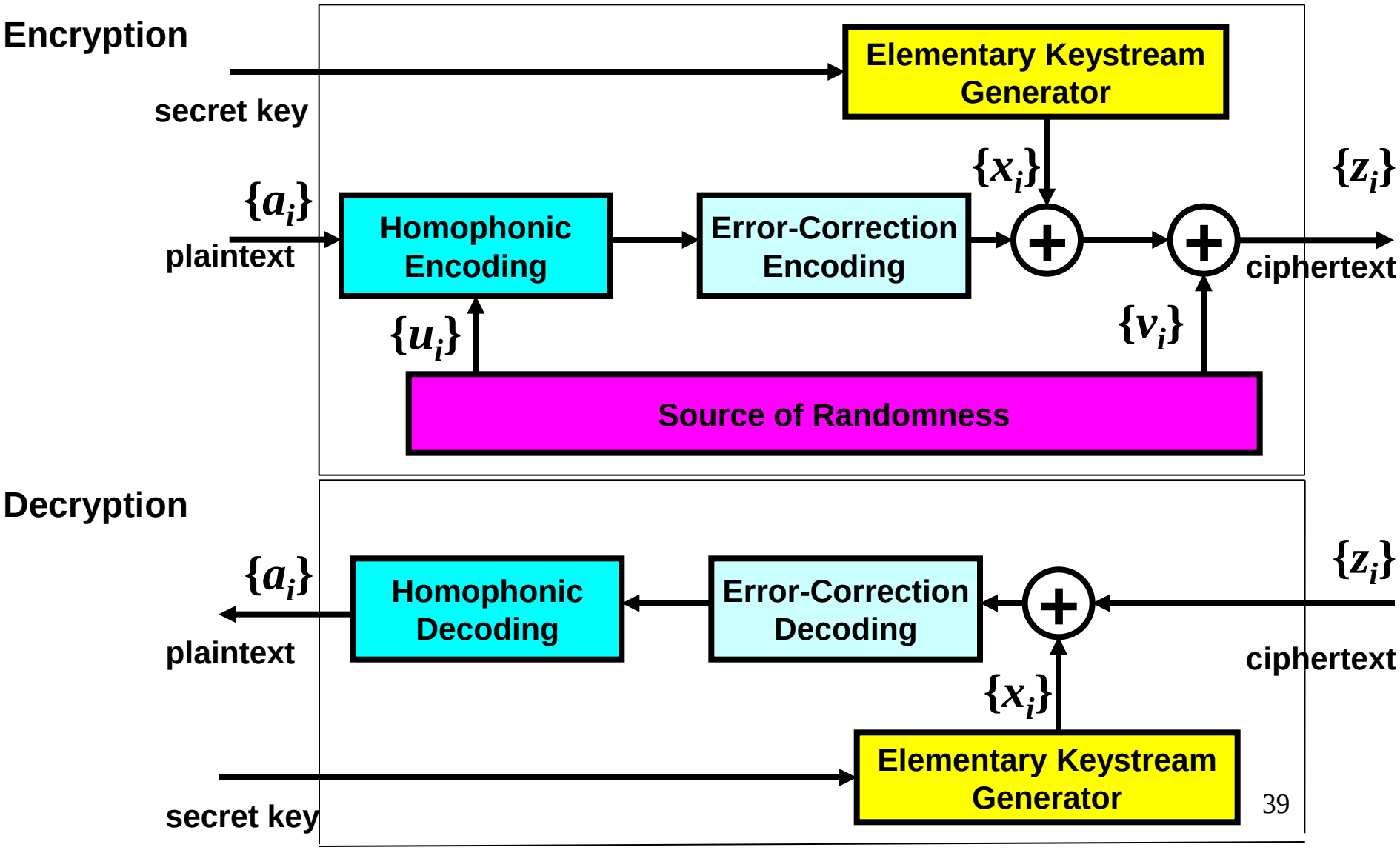
Effects:

- Enhanced Security Implied by Randomness
- Low Implementation Complexity

Homophonic Coding Based LPN Encryption



Framework for a Stream Ciphers Design



VI. Concluding Remarks

- The homophonic coding controlled by the randomness, provides that an attacker faces not only the traditional problems of cryptanalysis but also the problem of decoding without the secret key which appears as complex as the exhaustive search over the possible secret keys.

- The framework provides **computational-complexity security as hard as certain instantiations of the LPN problem.**
- Assuming availability of very short keystream segments only, the encryption framework provides certain level of **information-theoretic security**.

Thank You Very Much for the Attention,

and

QUESTIONS Please!