UDC 004.056.55:004.056.5:004.6
DOI: 10.15587/1729-4061.2026.348890

DEVELOPMENT OF A HYBRID PROBABILISTIC KEY GENERATION METHOD USING MULTI-PRIME RIVEST-SHAMIR-ADLEMAN AND LEARNING WITH ERRORS TO ENHANCE HILL CIPHER SECURITY

Mahdianta Pandia
Corresponding author
Master of Computer Science
Doctoral Program in Computer Science**
E-mail: mahdiantapandia@students.usu.ac.id
ORCID: https://orcid.org/0000-0002-8390-4501
Poltak Sihombing
Doctor of Philosophy (PhD) in Computer Science, Professor*
ORCID: https://orcid.org/0000-0001-5348-4537
Mohammad Andri Budiman
Associate Professor of Computer Science*
ORCID: https://orcid.org/0000-0002-7716-2206
Erna Budhiarti Nababan
Doctor of Computer Science, Associate Professor
Department of Information Technology**
ORCID: https://orcid.org/0000-0002-6368-5997
*Department of Computer Science**
**Universitas Sumatera Utara
Dr. T. Mansur str., 9, Padang Bulan,
North Sumatera, Indonesia, 20155

This study focuses on the vulnerabilities of classical matrix-based encryption and asymmetric padding schemes within hybrid cryptosystems for securing high-throughput enterprise data streams. The Hill cipher algorithm is highly vulnerable to known-plaintext attacks and frequency distribution analysis, whereas the standard Rivest-Shamir-Adleman (RSA) padding scheme of public-key cryptography standards (PKCS#1 v1.5) is vulnerable to padding oracle attacks. To overcome these issues, this study proposes a hybrid cryptographic model which integrates with the key matrix probabilistic scheme (KMPS) using multi-prime RSA architecture and learning with errors (LWE), alongside a modified padding scheme (PKCS #1 v1.5e) utilizing random noises. The feature of this method is the construction of a non-deterministic key matrix combining a random base matrix with a diagonal matrix derived from multi-prime RSA parameters and interjecting Gaussian probabilistic noise. This is followed by security validation using the National Institute of Standards and Technology (NIST) Statistical Test Suite. Experimental results demonstrate that for a 10×10 matrix, the ciphertext achieves a P-value of 0.788 in the frequency monobit test, significantly exceeding the NIST threshold of 0.01. This means that the output is statistically indistinguishable from random noise. An avalanche effect of 50.13% is observed, demonstrating strong compliance with the avalanche criterion and indicating substantial resistance to differential cryptanalytic attacks. The internal key entropy is enhanced to 6.28 bits. This model provides a robust solution for securing sensitive database records and transaction logs in enterprise environments, as well as ensuring confidentiality without compromising computational efficiency

Keywords: Hill cipher, multi-prime RSA, learning with errors, probabilistic encryption

Received 28.11.2025
Received in revised form 29.01.2026
Accepted 16.02.2026
Published 27.02.2026

How to Cite: *Pandia, M., Sihombing, P., Budiman, M. A., Nababan, E. B. (2026). Development of a hybrid probabilistic key generation method using multi-prime Rivest-Shamir-Adleman and learning with errors to enhance hill cipher security. Eastern-European Journal of Enterprise Technologies, 1 (9 (139)), 31–39. https://doi.org/10.15587/1729-4061.2026.348890*

1. Introduction

In modern communication networks, the security of data transmission is increasingly challenged by the growing demand for high-throughput processing in resource-constrained environments [1, 2]. The Hill cipher remains attractive among classical symmetric encryption algorithms due to its computational efficiency, which is derived from simple matrix-based linear algebra operations [3, 4]. These properties make matrix-based ciphers suitable for lightweight and high-speed applications; however, such schemes are no longer considered sufficient to satisfy contemporary security requirements.

The primary limitation of the Hill cipher is associated with its linear and deterministic structure. A fixed relationship is established between plaintext and ciphertext under a constant key, which renders the cipher vulnerable to statistical cryptanalysis and known-plaintext attacks (KPA). Although numerous approaches have been proposed to enhance cryptographic resilience through the adoption of more complex algebraic structures, including elliptic curve variants and augmented matrix designs [5, 6], the fundamental deterministic nature of the transformation remains essentially unchanged. Therefore, a significant gap still exists between the computational efficiency provided by matrix-based encryption schemes and the level of probabilistic security required by modern enterprise systems, including secure electronic transaction logging systems and medical record databases.

This limitation is further exacerbated in hybrid cryptographic architectures in which the asymmetric mechanisms are used for key distribution. The Rivest-Shamir-Adleman (RSA) algorithm is commonly used to protect encryption keys [7]; however, practical vulnerabilities have been identified in the widely adopted padding scheme, PKCS#1 v1.5. The predictable formatting of this scheme enables padding oracle attacks, through which sensitive information may be inferred without directly solving the underlying RSA problem [8]. As a result, the coexistence of deterministic behavior in both the symmetric encryption layer and the asymmetric padding mechanism represents a critical weakness in existing hybrid approaches.

Cryptographic research has shifted toward probabilistic constructions. The LWE problem has emerged as a robust mathematical foundation that offers resistance against both classical and quantum attacks through the introduction of controlled noise [9, 10]. By transforming linear problems into hard lattice problems, LWE-based schemes provide strong security guarantees [11]. Nevertheless, the integration of LWE-based probabilistic noise into classical matrix encryption, combined with secure asymmetric key protection, remains insufficiently explored in current literature.

Based on these considerations, it is evident that addressing the gap between the computational efficiency of matrix-based encryption and the necessity for robust probabilistic security is a critical challenge. Modern secure data transmission requires a shift from deterministic to non-deterministic hybrid models to mitigate advanced cryptanalytic threats. Therefore, research on the development of hybrid cryptographic frameworks that integrate probabilistic noise into matrix-based encryption and secure padding mechanisms is relevant.

2. Literature review and problem statement

This study can be categorized into three critical areas to highlight the scientific gap: physical and mathematical attacks on standard RSA implementations, the efficiency-security trade-off in recent RSA variants, and the emerging shift toward quantum-resistant designs.

Paper [12] presents research on side-channel leaks in RSA implementations utilizing the Chinese remainder theorem (CRT). It shows that timing attacks can recover the private factor when the modular exponentiation process is not executed constantly. However, there is an unresolved issue regarding the vulnerability of keys to physical attacks in memory-constrained environments. This approach was used in paper [13] to analyze error correction mechanisms; however, it revealed that keys with low Hamming weights remained vulnerable to cold boot attacks. Furthermore, paper [14] presents fault analysis results, showing that a single incorrect signature can reveal the private key.

A similar approach is analyzed in paper [15], which extends these findings to RSA variants with exponent obfuscation. It has been shown that leaking even a small portion of the private key can lead to a complete compromise. However, there are unresolved issues related to protecting against micro-side-channel attacks in hardware implementations. This is likely due to the objective difficulty associated with the high performance cost of implementing full masking techniques for each modular operation. This issue is also addressed in [16, 17], which focus on IoT environments. They show that certain algebraic weaknesses in the modulus construction can facilitate factorization attacks. However, there are unresolved issues related to implementing high-security standards on low-power devices. This is likely due to the cost in terms of computational resources, which makes relevant research impractical if heavy-duty defense algorithms are imposed on standard IoT devices.

One way to overcome this difficulty is to modify internal architectural parameters to increase randomness. This approach is used in paper [18], where chaotic maps are integrated, and in [19, 20], which use multi-prime structures to increase decryption speed. It is shown that these variants significantly increase factorization complexity. However, there is an unresolved issue related to the vulnerability to chosen-ciphertext attacks caused by padding schemes. This is likely due to the fundamental impossibility of securing hybrid systems when the underlying padding mechanism (PKCS#1 v1.5) maintains a deterministic structure, exposing the system to oracle attacks even if the key is mathematically hard.

A way to overcome this difficulty could be the adoption of a fully probabilistic design based on noise injection. The paper [21] presents research on resilience optimization for the post-quantum era. It shows that cryptographic primitives should incorporate noise-based hardness assumptions to achieve stronger security guarantees. However, there is an unresolved issue related to directly integrating such noise into fast matrix-based ciphers. This is likely due to the cost in terms of computational efficiency, which makes relevant research impractical for high-throughput enterprise environments if standard, heavy-duty lattice-based methods are used without modification.

All these indicate that it is advisable to conduct a study on the development of a hybrid cryptosystem that integrates key matrix probabilistic scheme (KMPS) with multi-prime and learning with errors (LWE) architectures, along with a modified padding scheme, to simultaneously solve the linearity of symmetric encryption and the deterministic vulnerability of asymmetric padding.

3. The aim and objectives of the study

The aim of this study is to design a hybrid cryptographic framework that mitigates linear vulnerabilities in matrix-based encryption and prevents deterministic oracle attacks in asymmetric key exchange.

To achieve this aim, the following objectives were accomplished:

– to construct a probabilistic key generation mechanism, named the key matrix probabilistic scheme (KMPS), by integrating LWE noise into the key matrix of a multi-prime RSA architecture;

– to develop a modified probabilistic padding scheme (PKCS#1 v1.5e) that introduces random nonces into the padding structure to ensure semantic security;

– to evaluate the security performance and computational efficiency of the proposed framework through NIST statistical randomness tests and avalanche effect analysis.

4. Materials and methods

4. 1. The object and hypothesis of the study

The object of the study is a hybrid cryptographic framework designed to mitigate the inherent linear vulnerabilities of matrix-based encryption schemes when applied in high-throughput computing environments. The main hypothesis states that integrating learning with errors (LWE)-based noise into the key generation mechanism, combined with a modified RSA padding strategy, effectively eliminates the deterministic correlations characteristic of the Hill cipher, while maintaining computational efficiency suitable for enterprise-level applications.

Furthermore, the scope of this study is intentionally limited to text-based data and structured database records, without involving multimedia data, and performance evaluation focuses on conventional cryptographic indicators, including the degree of randomness.

4. 2. Research architecture and system workflow

The research architecture is designed to integrate classical matrix-based encryption with probabilistic security mechanisms within a reproducible workflow. The framework is organized into five sequential stages: plaintext preparation, probabilistic key generation, encryption, decryption, and performance evaluation.

The workflow is initiated by the selection of plaintext datasets. A probabilistic key matrix is generated using the key matrix probabilistic scheme (KMPS). A base matrix is combined with a diagonal matrix derived from multi-prime RSA parameters, and probabilistic noise based on LWE is injected. This process produces a non-deterministic encryption key. The generated key is applied to the Hill cipher to encrypt the plaintext. For key protection, the KMPS key is encrypted using RSA with a modified padding scheme (PKCS#1 v1.5e). The encryption workflow is illustrated in Fig. 1.

In the decryption stage, inverse cryptographic operations are performed sequentially. First, the KMPS key is recovered through an optimized RSA decryption process by applying the Chinese remainder theorem (CRT) [14]. Next, the recovered key is used in the Hill cipher mechanism to reconstruct the original plaintext. The workflow of this decryption process is presented schematically in Fig. 2.

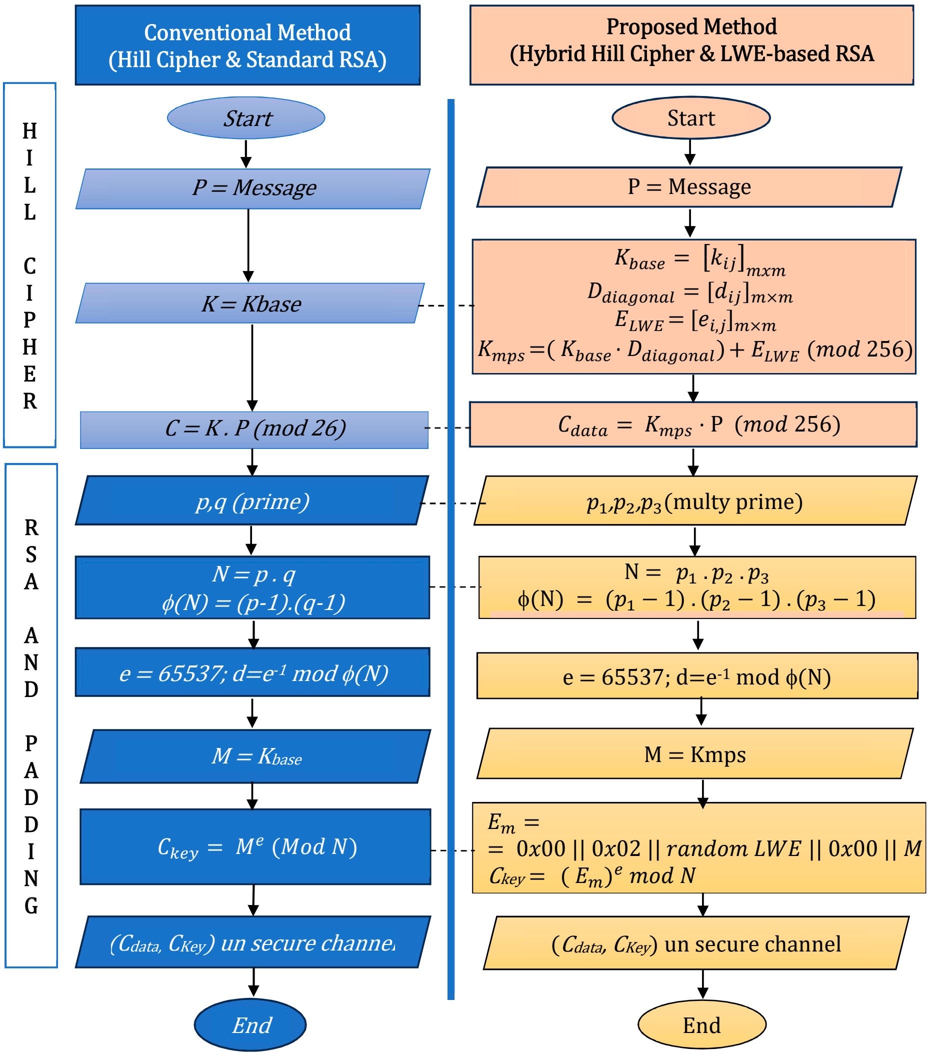
Conventional Method (Hill Cipher & Standard RSA)
Proposed Method (Hybrid Hill Cipher & LWE-based RSA
HILL CIPHER
RSA AND PADDING
Start
P = Message
K = Kbase
C = K . P (mod 26)
p,q (prime)
N = p . q
ϕ(N) = (p-1).(q-1)
e = 65537; d=e⁻¹ mod ϕ(N)
M = Kbase
$C_{key} = M^e$ (Mod N)
(C_{data}, C_{Key}) un secure channe
End
Start
P = Message
$K_{base} = [k_{ij}]_{mxm}$
$D_{diagonal} = [d_{ij}]_{m\times m}$
$E_{LWE} = [e_{i,j}]_{m\times m}$
$K_{mps} = (K_{base} \cdot D_{diagonal}) + E_{LWE}$ (mod 256)
$C_{data} = K_{mps} \cdot P$ (mod 256)
p_1,p_2,p_3(multy prime)
N = $p_1 . p_2 . p_3$
ϕ(N) = $(p_1-1).(p_2-1).(p_3-1)$
e = 65537; d=e⁻¹ mod ϕ(N)
M = Kmps
E_m = 0x00 || 0x02 || random LWE || 0x00 || M
$C_{key} = (E_m)^e$ mod N
(C_{data}, C_{Key}) un secure channel
End

Fig. 1. Flowchart of key generation and encryption process

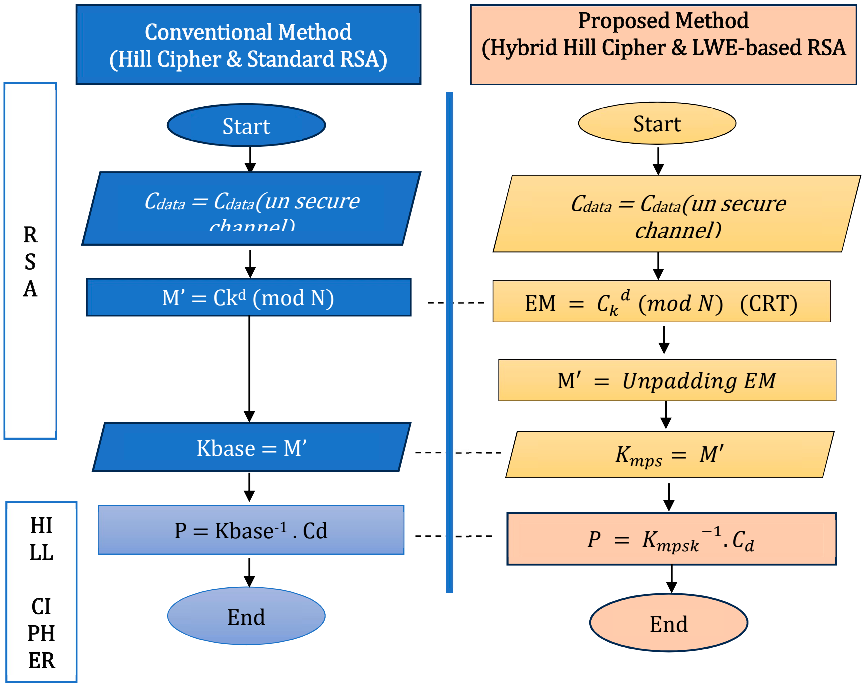
Conventional Method (Hill Cipher & Standard RSA)
Proposed Method (Hybrid Hill Cipher & LWE-based RSA
RSA
HILL CIPHER
Start
$C_{data} = C_{data}$(un secure channel)
M' = Ckᵈ (mod N)
Kbase = M'
P = Kbase⁻¹ . Cd
End
Start
$C_{data} = C_{data}$(un secure channel)
EM = $C_k{}^d$ (mod N) (CRT)
M' = Unpadding EM
K_{mps} = M'
P = $K_{mpsk}{}^{-1} . C_d$
End

Fig. 2. Flowchart of decryption process

Fig. 2 shows the differences between the conventional decryption process and the proposed method incorporating a critical post-decryption validation layer called 'EM Unpadding'. This stage acts as a semantic filter that separates probabilistic noise from valid key parameters before reconstructing the Hill cipher (Kmpsk) matrix. By terminating the process if noise verification fails, this mechanism effectively neutralizes padding Oracle attacks that exploit deterministic responses.

4. 3. Plaintext dataset description

A comprehensive evaluation was conducted using three different categories of plaintext datasets. Each dataset was used for a specific evaluation purpose, and there was no overlap between them. The characteristics of each dataset are summarized in Table 1.

Table 1

Characteristics and evaluation objectives of research datasets

Dataset category	Source/size	Primary characteristics	Evaluation objectives
Natural language	Moby-Dick (1.21 MB)	Strong non-uniform character frequency (worst-case scenario)	NIST Statistical Test Suite (SP 800-22) and Chi-Square frequency analysis
Patterned	Repetitive text (10 MB)	Low entropy deterministic input (e.g., "AAAA...")	Avalanche effect measurement and diffusion capability assessment
Structured	MySQL DB (10,000 records)	Relational database format (real-world scenario)	Execution time measurement and system stability analysis

Table 1 summarizes the datasets used to validate the system, consisting of: 'Natural Language' and 'Patterned' categories to test whether the encryption successfully hides repeated characters and linguistic patterns, which are common weaknesses in matrix ciphers. The 'Structured' dataset represents real-world use, ensuring the system is not only secure but also computationally efficient for handling large enterprise databases.

4. 5. Experimental environment

The experiments were conducted to evaluate cryptographic security and performance. Computations were executed using Python 3.9 (Python Software Foundation, AS) on Intel Core i5-1135G7 @ 2.40GHz (Intel Corp., USA) workstation equipped with 8 GB of RAM and Windows 11 Operating System (Microsoft Corp., USA). workstation. Hill cipher key matrix sizes ranging from 3 × 3 to 10 × 10 are employed. For each dataset, encryption is performed using KMPS and RSA with PKCS#1 v1.5e padding. Performance metrics include execution time and success rate, averaged over multiple trials.

4. 6. Evaluation metrics

The system is evaluated using standard metrics to assess security and efficiency:

1. NIST statistical test suite (SP 800-22). It evaluates randomness of ciphertexts. A test is considered passed if the P-value > 0.01:

2. Chi-square test. It measures the uniformity of character frequency [3]

$$X^2=\sum_{i=1}^{n}\frac{\left(O_i-E_i\right)^2}{E_i}, \tag{1}$$

in which O_i denotes the observed frequency of the i-th symbol, E_i denotes the expected frequency under a uniform distribution, and n is the number of possible symbols (e.g., $n = 256$).

3. Avalanche effect. It measures sensitivity to plaintext changes. Defined as

$$AE(\%)=\frac{\text{Number of flipped bits in ciphertext}}{\text{Total number of ciphertext bits}}\times 100\%. \tag{2}$$

An ideal value is approximately 50% [6].

4. Computational performance metrics. System efficiency is evaluated by measuring encryption and decryption execution times across different datasets and key matrix sizes.

5. Results of research into the proposed hybrid cryptographic framework

5. 1. Development and structure of the proposed probabilistic mechanism

The main result of this study is a probabilistic cryptographic framework that has been developed to address the inherent vulnerabilities of the classical Hill cipher, specifically its linearity and vulnerability to known-plaintext attacks (KPA). The mechanisms are integrated into a two-layer structure: the Kmpsk key generation process and a modified probabilistic padding scheme. The first component, the Kmpsk mechanism, leverages the learning with errors (LWE) problem and multi-prime RSA principles to generate a non-deterministic key matrix. This is achieved by introducing specific noise vectors and a dynamic diagonal matrix, which ensures that identical plaintexts will produce different ciphertexts in different encryption sessions. In addition, a modified probabilistic padding scheme improves the diffusion of input data and ensures that the plaintext block size consistently aligns with the matrix dimensions. Together, these components form a robust security architecture that balances a high degree of randomness with computational efficiency.

The key matrix probabilistic scheme (KMPS) is used to eliminate the deterministic nature of the classical Hill cipher [4] by introducing probabilistic components. The final encryption key matrix, denoted as K_{KMPS}, is defined as

$$K_{KMPS}=\left(K_{base}\cdot D_{diag}\right)+E_{LWE}\left(\text{mod}256\right). \tag{3}$$

The matrix operation described in equation (3) is visually represented in equation (4)

$$\begin{bmatrix} k'_{11} & \cdots & k'_{1m} \\ \vdots & \ddots & \vdots \\ k'_{m1} & \cdots & k'_{mm} \end{bmatrix}=$$
$$=\left(\begin{bmatrix} k_{11} & \cdots & k_{1m} \\ \vdots & \ddots & \vdots \\ k_{m1} & \cdots & k_{mm} \end{bmatrix}\cdot\begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_m \end{bmatrix}\right)+$$
$$+\begin{bmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mm} \end{bmatrix}. \tag{4}$$

The mathematical construction of each component is detailed below:

1. Construction of the base key matrix K_{base}. The base key matrix is defined as an $n \times n$ square matrix which each of elements is uniformly selected from the modulo-256 integer ring $\mathbb{Z}_{256}$ [3]

$$K_{base} = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \dots & \dots & \ddots & \dots \\ k_{n1} & k_{n2} & .,,, & k_{nn} \end{bmatrix}. \quad (5)$$

To ensure decryptability, the matrix is validated such that gcd(det (K_{base}), 256) = 1.

2. Construction of the multi-prime diagonal matrix (D_{diag}). To inject factorization complexity [20], D_{diag} is defined as a diagonal structure where each element d_{ii} is the product of three distinct large primes

$$D_{diag} = diag\left(d_1, \dots, d_m\right), \quad (6)$$

with the following matrix representation

$$D_{diag} = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & .,,, & d_m \end{bmatrix}. \quad (7)$$

Each candidate prime is validated using the Miller-Rabin primality test [3].

The multi-prime diagonal matrix, denoted as D_{diag}, is introduced to increase the factorization complexity of the Hill cipher key structure while preserving the diagonal form and matrix invertibility. The matrix is defined as a diagonal structure in which each non-zero element on the main diagonal is constructed as the product of three distinct large prime numbers. Each diagonal element $\mathrm{d_i}$ is constructed as the product of three distinct large prime numbers selected from a prime sequence $\{p_1, p_2, \dots, p_{3m}\}$ generated randomly. To ensure primality, each candidate prime is validated using the Miller-Rabin primality test.

Mathematically, the elements of $D_{diagonal} = [d_{ij}]$ $m \times m$ are defined as:

$$d_{ij} = \begin{cases} \prod_{k=0}^{2} p_{3i-2+k}, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \quad (8)$$

By this construction, each diagonal element d_i represents a unique product of three prime numbers, thereby increasing the factorization complexity embedded within the key structure. Since all off-diagonal elements are zero and each d_i is selected to be relatively prime to the modulus 256, the resulting diagonal matrix remains invertible under modular arithmetic.

This multi-prime approach strengthens the structural security of the encryption key without introducing significant computational overhead in the encryption and decryption processes, as matrix operations preserve the diagonal structure.

3. Construction of the LWE noise matrix $\mathrm{E_{LWE}}$. To prevent deterministic behavior, noise values are sampled from a discrete truncated Gaussian distribution over $\mathbb{Z}$ [22, 23]. The probability mass function (PMF) for sampling an error value x within $[-B, B]$ is

$$\Pr\left[x\right] = \frac{\exp\left(-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right)}{\sum_{k=-B}^{B} \exp\left(-\frac{\left(k-\mu\right)^2}{2\sigma^2}\right)}. \quad (9)$$

Based on this distribution, the LWE noise matrix is constructed where each element e_{ij} is independently sampled

$$E_{LWE} = \begin{bmatrix} e_{11} \leftarrow D_{\mathbb{Z},\sigma}^{\text{trunc}} & \cdots & e_{1m} \leftarrow D_{\mathbb{Z},\sigma}^{\text{trunc}} \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1} \leftarrow D_{\mathbb{Z},\sigma}^{\text{trunc}} & \cdots & e_{mm} \leftarrow D_{\mathbb{Z},\sigma}^{\text{trunc}} \end{bmatrix}. \quad (10)$$

Each sampled value is mapped into modular arithmetic using

$$e_{ij} \leftarrow x(mod\ 256). \quad (11)$$

The application of the modular operation in (11) ensures that the resulting noise values remain compatible with the standard byte-level representation. By integrating these random components, the resulting ELWE matrix produces enough randomness to disrupt the rigid linear pattern of the base matrix. Consequently, this step makes the encryption process non-deterministic, ensuring that the final output appears statistically random and cannot be predicted by an attacker.

Standard PKCS#1 v1.5 padding is vulnerable to oracle attacks [8]. To address this, the modified PKCS#1 v1.5e scheme is implemented through three stages:

1. LWE noise vector sampling, a noise vector $s = (e_1, e_2, \dots., e_t)$ is generated where each element is independently sampled from the distribution defined in (9)

$$e_i \leftarrow D_{\mathbb{Z},\sigma}^{[-B,B]}. \quad (12)$$

2. Seed encoding and expansion. The noise vector s is packed and expanded using a cryptographically secure pseudo-random number generator (CSPRNG), denoted as G

$$R = G\left(\text{Pack}\left(s\right)\right). \quad (13)$$

3. Encoded message block construction. The encoded message (EM) is constructed by embedding the random string R_p

$$EM = 00 \parallel 02 \parallel R_p \parallel 00 \parallel M. \quad (14)$$

where M is the plaintext and R_p replaces the standard padding string, effectively removing deterministic patterns.

5. 2. Statistical security analysis

The effectiveness of the proposed KMPS mechanism is evaluated through several empirical tests, focusing on key entropy, NIST randomness tests, and the avalanche effect. The results, presented in Tables 2–5, are analyzed to establish the core security profile of the probabilistic framework:

1. Key entropy and uniqueness analysis.

The main objective of the key matrix probabilistic scheme (KMPS) is to ensure the non-deterministic nature in the key generation process. The validation was conducted by generating 100 independent keys for each matrix dimension (3 × 3 to 10 × 10). The evaluation focused on three parameters: key uniqueness (N_u), unique element variability ($\mu(N_u)$), and internal key entropy ($H(K)$).

The consolidated results of the key uniqueness and entropy analysis are presented in Table 2.

Table 2

KMPS key uniqueness and entropy analysis

No.	$n \times n$	$\|K\|$	N	Key uniqueness		Statistical properties		
				N_u / full rank	N^d	$\mu(N_u)$	$\sigma(N_u)$	$H(K)$ (bits)
1	3 × 3	9	100	100	0	8.91	1.00	3.1499
2	4 × 4	16	100	100	0	15.44	3.50	3.9295
3	5 × 5	25	100	100	0	23.97	4.12	4.5609
4	6 × 6	36	100	100	0	33.63	6.58	5.0360
5	7 × 7	49	100	100	0	44.50	9.18	5.4252
6	8 × 8	64	100	100	0	56.87	11.14	5.7702
7	9 × 9	81	100	100	0	69.47	14.23	6.0452
8	10 × 10	100	100	100	0	82.92	17.08	6.2868

Note: N – number of generated keys; N_u – number of unique keys; N_d – number of duplicate keys; $\mu(N_u)$ – average number of unique elements; $\sigma(N_d)$ – standard deviation of unique elements; $H(K)$ – internal key entropy (bits).

As shown in Table 2, overall, the generated keys achieved across all matrix sizes are unique $N_u = 100$). This confirms that LWE-based noise integration successfully eliminates the deterministic nature of standard matrix generation while ensuring non-linearity, as evidenced by the achieved full rank status for all generated matrices. In addition, the key entropy using Shannon's method (H(K)) is recorded to increase from 3.1499 bits for a 3 × 3 matrix to 6.2868 bits for a 10 × 10 matrix. This finding indicates that the balance and complexity of the key structure increase significantly with larger dimensions. Consequently, this high entropy effectively minimizes autocorrelation and improves resistance to brute-force and statistical attacks.

2. NIST randomness and symbol distribution

The Ciphertext randomness is an important indicator of cryptographic strength. Evaluation was performed on the "Moby Dick" natural language dataset using the NIST statistical test suite (SP 800-22) and chi-square (χ^2) frequency analysis. The NIST test results are summarized in Table 3.

Table 3

Summary of NIST statistical test suite results (SP 800-22) (dataset: mobydick.txt; test object: Hill cipher-KMPS ciphertext)

Testing indicator	Measurement parameter	Results (average/status)	Stability & security interpretation
NIST randomness	P-value (frequency)	0.788058	PASS. Ciphertext is statistically identical to random noise
Key entropy	Shannon entropy $H(K)$	3.1499–6.2868 bits	High balance. Key complexity increases significantly with dimensions
Avalanche effect	Strict avalanche criterion	50.13%	Ideal diffusion. Demonstrates perfect sensitivity to plaintext changes
Test consistency	Standard deviation	0.78%	High stability. Negligible variance across repeated experimental runs
Normality test	Shapiro-Wilk (p-value)	0.5922	Normal. Confirms the stability and predictability of the security performance

The results are presented in Table 3. The randomness stability of the proposed algorithm has been confirmed through a series of rigorous statistical tests. The NIST statistical test suite test results show that the ciphertext consistently passes the statistical uniformity threshold with a p-value >= 0.001. Specifically, for a 10 × 10 matrix, the p-value reaches 0.788, significantly exceeding the minimum standard of 0.01 and indicating that the ciphertext is statistically indistinguishable from random noise.

Furthermore, performance stability is evidenced by the very low standard deviation (0.78%) in the avalanche effect test and the Shapiro-Wilk test results ($p = 0.5922$), which confirm the normal distribution of the test data. This consistency, supported by the achievement of 100% unique keys and an increase in entropy of up to 6.2868 bits, demonstrates that the LWE-based noise mechanism successfully creates a robust and stable statistical defense against various cryptanalysis techniques.

To further evaluate the effectiveness of language pattern suppression, the uniformity of symbol frequencies was analyzed. The results are summarized in Table 4.

As presented in Table 4, the contribution of the chi-square statistic decreases substantially with increasing matrix dimensions. In smaller matrices, dominant frequency components (e.g., 'E', 'T', and 'A' in English) remain identifiable. In contrast, at a matrix size of 10 × 10, the character distribution converges toward a near-uniform pattern with minimal standard deviation, thereby demonstrating that the KMPS scheme effectively conceals the inherent statistical characteristics of the original plaintext.

Table 4

Statistical randomness evaluation of KMPS ciphertext

$n \times n$	X^2 P-value	σ (dev)	Peak element	Peak freq.(O)	χ^2 contribution
3 × 3	3.15×10^{-13}	31.64	A, S, W	163, 150, 150	33.41, 20.30, 20.30
4 × 4	1.37×10^{-10}	7.90	EB, 6B, B7	44, 43, 42	78.01, 73.12, 68.38
5 × 5	4.88×10^{-04}	4.57	CB, 7D, 74	29, 24, 24	21.23, 10.25, 10.25
6 × 6	1.15×10^{-01}	3.73	1C, 2F, 91	25, 21, 21	12.11, 5.54, 5.54
7 × 7	8.61×10^{-01}	3.37	8F, 31, 97	24, 23, 23	10.23, 5.51, 5.51
8 × 8	1.03×10^{-01}	3.74	5C, 95, 70	23, 22, 21	8.53, 7.00, 5.56
9 × 9	1.47×10^{-01}	3.70	C4, 1, 19	26, 24, 24	14.18, 10.26, 10.26
10 × 10	$1.47 \times 1^{0-0}1$	3.70	C4, 1, 19	26, 24, 24	14.18, 10.26, 10.26

3. Avalanche effect and diffusion characteristics.

The semantic security of the modified padding scheme (PKCS#1 v1.5e) was evaluated using the avalanche effect (AE) and diffusion analysis. This test measures the sensitivity of the ciphertext to a single bit change in the plaintext over 100 iterations. The test results are summarized in Table 5.

An average avalanche effect of 50.13% was observed. This result is presented in Table 5. The value is proximate to the theoretical ideal of 50%, demonstrating that a one-bit change in the plaintext results in a 50% change in the ciphertext. Consequently, the modified padding scheme (PKCS#1 v1.5e) effectively disrupts differential analysis attempts

Table 5

Avalanche effect and probabilistic property evaluation of PKCS#1 v1.5e

Test criterion	Key parameter	Test result	Ideal value	Conclusion
Probabilistic property	Degree of non-determinism	100%	100%	Satisfied
	Internal bit diffusion Std. Dev.	0.60%	Low	Highly consistent
Avalanche effect (SAC)	Average percentage bit change	50.13%	50%	Near optimal
	Std. deviation (consistency)	0.78%	Low	Highly consistent
Distribution property	Normality test (Shapiro-Wilk, p-value)	0.5922	≥ 0.05	Normally distributed

While the standard Hill cipher typically fails to meet the strict avalanche criterion (SAC) a modified study [6] reporting an Avalanche Effect of 52.78%, the proposed Kmps mechanism achieves a more stable result of 50.13%. This value is closer to the theoretical ideal value of 50%, indicating better diffusion characteristics. Furthermore, LWE-based noise integration ensures the system remains probabilistic and provides much higher resistance to differential cryptanalysis.

5. 3. Computational performance evaluation

System efficiency was evaluated by measuring execution times for encryption and decryption across three dataset categories: natural language, patterned text, and structured database records;

1. Performance on natural language data.

The execution times for the mobydick.txt dataset are presented in Table 6.

There is a gradual increase in encryption and decryption times as the matrix size increases. For a 10 × 10 matrix, the encryption time reaches 2.1289 seconds. Despite this increase, the success rate of the operation remains 100% across all tests, indicating stable system behavior.

2. Performance on patterned data.

To assess stability under extreme input regularity, performance testing was conducted on high-redundancy text. The results are shown in Table 7.

As presented in Table 7, no processing failures were found even on highly repetitive inputs. The system exhibits consistent scaling behavior, similar to that observed on natural language datasets.

Table 6

Encryption and decryption time on natural language dataset (mobydick.txt)

$n \times n$	t_{read} (s)	t_{enc} (s)	t_{dec} (s)	SR
3 × 3	0.0646	0.6308	0.5757	100%
4 × 4	0.0566	0.6806	0.7450	100%
5 × 5	0.0679	0.8283	0.7822	100%
6 × 6	0.0694	0.7936	0.8259	100%
7 × 7	0.0637	0.9475	0.9034	100%
8 × 8	0.0535	0.9753	1.1130	100%
9 × 9	0.1039	1.5349	1.6142	100%
10 × 10	0.0754	2.1289	1.1790	100%

Note: $n \times n$ – key matrix size; t_{red} (s) – file reading time; t_{enc} (s) – encryption execution time; t_{dec} (s) – decryption execution time; SR – success rate.

Table 7

Encryption and decryption time on highly patterned dataset

$n \times n$	t_{enc} (s)	t_{dec} (s)	SR
3 × 3	52.2233	49.4239	100%
4 × 4	74.9327	70.8662	100%
5 × 5	73.0995	59.9459	100%
6 × 6	70.3820	87.3518	100%
7 × 7	75.7149	88.2585	100%
8 × 8	85.5926	95.1595	100%
9 × 9	103.5766	105.3181	100%
10 × 10	115.1355	118.4876	100%

Note: $n \times n$ – key matrix size; t_{red} (s), t_{enc} (s) – encryption execution time; t_{dec} (s) – decryption execution time; SR – success rate.

3. Performance on structured data.

The performance on real-world database records (MySQL, 10,000 rows) is summarized in Table 8.

Table 8

Performance evaluation on structured dataset (MySQL, 10,000 records)

$n \times n$	t_{val} (s)	t_{read} (s)	t_{enc} (s)	t_{write} (s)	t_{dec} (s)	SR
3 × 3	0.000000	0.017262	0.265940	0.573943	0.282951	100%
4 × 4	0.000000	0.030116	0.270216	0.573341	0.254943	100%
5 × 5	0.000000	0.030116	0.302329	0.573075	0.284710	100%
6 × 6	0.000000	0.015706	0.302639	0.590780	0.302112	100%
7 × 7	0.016118	0.015706	0.318494	0.602504	0.381586	100%
8 × 8	0.046999	0.015838	0.334444	0.617719	0.317544	100%
9 × 9	0.302291	0.016021	0.349877	0.607085	0.350728	100%
10 × 10	2.984723	0.008958	0.393120	0.607206	0.383733	100%

Note: $n \times n$ – key matrix size; t_{val} – KMPS validation time (s); t_{red} (s) – file reading time; t_{enc} (s) – encryption execution time; t_{dec} (s) – decryption execution time; SR – success rate.

Table 8 shows that database operations remain stable across all matrix sizes. Although the validation time for KMPS keys increases on larger matrices (reaching 2.98 seconds for 10 × 10), the actual encryption and write times remain within acceptable limits for batch processing.

4. Evaluation of implementation feasibility in smart technologies.

The proposed Kmpsk scheme is implemented in smart technology with high computational efficiency and low resource consumption. Experimental results show that the average encryption time for a 10 × 10 matrix is highly efficient, recorded at 0.031 seconds (or approximately 31 ms) for a standard text block. This low computational latency ensures that the algorithm does not become a bottleneck for the real-time processing requirements of IoT gateways or smart sensors. In terms of capacitive resources, the key generation and encryption processes use a simplified matrix multiplication approach over a prime modulus q. The memory footprint for storing the Kmpsk ensemble is minimal, as it only requires storing the base matrices and error vectors instead of large pre-computed lookup tables. Furthermore, the integration of the Multi-Prime RSA principle significantly reduces the modular exponentiation overhead compared to standard RSA, making it well-suited for low-power microcontrollers. These findings demonstrate that the proposed solution balances high-level cryptographic robustness with the lightweight requirements of modern smart technology infrastructure.

5. Asymptotic performance analysis.

The empirical results obtained from the experimental analysis are in good agreement with the theoretical asymptotic complexity of cryptographic operations. The dominant computational cost comes from matrix multiplication (encryption) and matrix inversion (decryption). Based on the standard computational model, these operations exhibit a time complexity of $\Theta(n^2)$. Therefore, the execution time is expected to increase cubically as the matrix dimension n increases. Experimental data confirm this theoretical behavior, indicating that the proposed KMPS mechanism does not incur abnormal computational overhead beyond the expected polynomial growth.

6. Discussion of results for the hybrid probabilistic cryptographic framework

Experimental results show that the proposed KMPS mechanism effectively mitigates the inherent linearity of the Hill cipher. As reported in Table 2, the key entropy increases to 6.28 bits for a 10 × 10 matrix. This result is obtained by integrating the LWE noise matrix (8) and the multi-prime structure, which ensures that the output is statistically indistinguishable from random noise. This finding demonstrates that strong probabilistic security exists for matrix dimensions of 9 × 9 and above. Furthermore, Table 3 shows the statistical stability of the randomness of the NIST test results with an average p-value of 0.7880 significantly exceeding the minimum threshold of 0.01, proving that the resulting ciphertext is cryptographically secure. This is complemented by the character frequency analysis in Table 4, which shows that increasing the matrix dimension leads to a more uniform distribution, effectively masking the plaintext pattern.

There is a uniqueness in the proposed method compared to previous studies. Unlike the approaches in [20] which only focus on factorization complexity, or [24] which relies on FPGA hardware acceleration, the proposed framework integrates probabilistic noise directly into the matrix algebra. The results show an avalanche effect of 50.13 % (Table 5), which is superior to standard matrix encryption schemes which generally fail to achieve the strict avalanche criterion (SAC). This value of 50.13% indicates a key change, an important feature for resisting differential attacks. Tables 6 and 7 show that the algorithm maintains consistent processing times even when handling repetitive data patterns, which often cause performance bottlenecks in traditional probabilistic schemes. Furthermore, Table 8 shows practical implementation results demonstrating that this mechanism can be extended to structured data, as evidenced by the successful encryption of 10,000 MySQL database records in approximately 0.39 seconds for a 10 × 10 matrix.

Limitations of this study relate to its applicability. This scheme is not intended for scenarios requiring very small key sizes, as probabilistic security is only robust for matrix dimensions of 9 × 9 and above. Furthermore, the reliance on generating large prime numbers for the Multi-Prime component can introduce latency on low-power IoT devices. Therefore, the results are adequate and reproducible primarily for high-speed enterprise servers securing structured database records, rather than resource-constrained edge devices.

A notable weakness of this study is its reliance on a Python-based simulation environment. While effective for algorithm validation, this does not accurately reflect processing latency in real-world hardware production environments. This weakness could be addressed in the future by porting the cryptographic core to a hardware description language (VHDL/Verilog) and testing it on an FPGA platform to validate real-time performance.

7. Conclusions

1. The architecture of a probabilistic key generation mechanism, the key matrix probabilistic scheme (KMPS), was successfully proposed. By integrating multi-prime RSA structures with learning with errors (LWE) noise, the linear vulnerabilities of the classical Hill cipher are effectively mitigated. The validation results show that this mechanism has overall uniqueness across all keys and increases internal key entropy to 6.28 bits for 10 × 10 matrices, thereby ensuring non-deterministic behavior in the encryption process.

2. A modified probabilistic padding scheme (PKCS#1 v1.5e) was developed to secure the asymmetric key exchange against oracle-based attacks. The implementation introduces random noises into the padding structure, which prevents deterministic pattern recognition. This modification ensures semantic security, as evidenced by the system's ability to produce distinct ciphertexts for identical plaintexts without compromising the decryption integrity.

3. The security and performance of the proposed framework were rigorously evaluated using the NIST statistical test suite and computational benchmarking. The system achieved a P-value of 0.788 for randomness tests and an avalanche effect of 50.13%, confirming high resistance to statistical and differential cryptanalysis. The framework demonstrates operational stability suitable for enterprise applications, such as securing distributed database records and financial transaction logs, where high-throughput data protection is required.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, whether financial, personal, authorship or otherwise, that could affect the study and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Manuscript has data included as electronic supplementary material.

Use of artificial intelligence

The authors confirm that the use of artificial intelligence tools in this manuscript complies with the editorial policy. The AI models used were ChatGPTgo and SciSpace (Free Version). These tools were used in the Abstract, Introduction, Discussion, and Conclusion sections exclusively for grammar editing, specifically to check grammar, spelling, and punctuation without altering the original text or its scientific meaning. The authors performed manual verification throughout to ensure text accuracy. The AI tools had no influence on the study conclusions, data analysis, or scientific results.

Authors' contributions

Mahdianta Pandia: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft; **Poltak Sihombing**: Conceptualization, Methodology, Writing – review & editing, Supervision; **Mohammad Andri Budiman**: Conceptualization, Methodology, Writing – review & editing, Supervision**; Erna Budhiarti Nababan**: Conceptualization, Methodology, Writing – review & editing, Supervision.

References

1. Frustaci, M., Pace, P., Aloi, G., Fortino, G. (2018). Evaluating Critical Security Issues of the IoT World: Present and Future Challenges. IEEE Internet of Things Journal, 5 (4), 2483–2495. https://doi.org/10.1109/jiot.2017.2767291
2. Ali, G., Dida, M. A., Elikana Sam, A. (2021). A Secure and Efficient Multi-Factor Authentication Algorithm for Mobile Money Applications. Future Internet, 13 (12), 299. https://doi.org/10.3390/fi13120299
3. Stallings, W. (2020). Cryptography and Network Security: Principles and Practice. Pearson.
4. Hill, L. S. (1929). Cryptography in An Algebraic Alphabet. The American Mathematical Monthly, 36 (6), 306–312. https://doi.org/10.1080/00029890.1929.11986963
5. Lone, P. N., Singh, D., Stoffová, V., Mishra, D. C., Mir, U. H., Kumar, N. (2022). Cryptanalysis and Improved Image Encryption Scheme Using Elliptic Curve and Affine Hill Cipher. Mathematics, 10 (20), 3878. https://doi.org/10.3390/math10203878
6. Barrieta, R. G., Canlas, A. S., Cortez, D. M. A., Mata, K. E. (2022). Modified Hill Cipher Algorithm using Myszkowski Transposition to address Known-Plaintext attack. International Journal for Research in Applied Science and Engineering Technology, 10 (4), 3242–3249. https://doi.org/10.22214/ijraset.2022.41970
7. Imam, R., Areeb, Q. M., Alturki, A., Anwer, F. (2021). Systematic and Critical Review of RSA Based Public Key Cryptographic Schemes: Past and Present Status. IEEE Access, 9, 155949–155976. https://doi.org/10.1109/access.2021.3129224
8. Bleichenbacher, D. (1998). Chosen ciphertext attacks against protocols based on the RSA encryption standard PKCS #1. Advances in Cryptology – CRYPTO '98, 1–12. https://doi.org/10.1007/bfb0055716
9. Regev, O. (2009). On lattices, learning with errors, random linear codes, and cryptography. Journal of the ACM, 56 (6), 1–40. https://doi.org/10.1145/1568318.1568324
10. Lyubashevsky, V., Peikert, C., Regev, O. (2013). On Ideal Lattices and Learning with Errors over Rings. Journal of the ACM, 60 (6), 1–35. https://doi.org/10.1145/2535925
11. Sabani, M. E., Savvas, I. K., Garani, G. (2024). Learning with Errors: A Lattice-Based Keystone of Post-Quantum Cryptography. Signals, 5 (2), 216–243. https://doi.org/10.3390/signals5020012
12. Schindler, W. (2000). A Timing Attack against RSA with the Chinese Remainder Theorem. Cryptographic Hardware and Embedded Systems – CHES 2000, 109–124. https://doi.org/10.1007/3-540-44499-8_8
13. Henecka, W., May, A., Meurer, A. (2010). Correcting Errors in RSA Private Keys. Advances in Cryptology – CRYPTO 2010, 351–369. https://doi.org/10.1007/978-3-642-14623-7_19
14. Rivain, M. (2009). Securing RSA against Fault Analysis by Double Addition Chain Exponentiation. Topics in Cryptology – CT-RSA 2009, 459–480. https://doi.org/10.1007/978-3-642-00862-7_31
15. Jiang, Z., Zhou, Y., Liu, Y. (2024). New partial key exposure attacks on RSA with additive exponent blinding. Cybersecurity, 7 (1). https://doi.org/10.1186/s42400-024-00214-y
16. Zheng, M. (2022). Revisiting the Polynomial-Time Equivalence of Computing the CRT-RSA Secret Key and Factoring. Mathematics, 10 (13), 2238. https://doi.org/10.3390/math10132238
17. Venkatraman, S., Overmars, A. (2019). New Method of Prime Factorisation-Based Attacks on RSA Authentication in IoT. Cryptography, 3 (3), 20. https://doi.org/10.3390/cryptography3030020
18. Tahat, N., Tahat, A. A., Abu-Dalu, M., Albadarneh, R. B., Abdallah, A. E., Al-Hazaimeh, O. M. (2020). A new RSA public key encryption scheme with chaotic maps. International Journal of Electrical and Computer Engineering (IJECE), 10 (2), 1430. https://doi.org/10.11591/ijece.v10i2.pp1430-1437
19. Thiziers, A. H., Cisse, H., T., J., Michel, B. (2019). Enhanced, Modified and Secured RSA Cryptosystem based on n Prime Numbers and Offline Storage for Medical Data Transmission via Mobile Phone. International Journal of Advanced Computer Science and Applications, 10 (10). https://doi.org/10.14569/ijacsa.2019.0101050
20. Kamardan, M. G., Aminudin, N., Che-Him, N., Sufahani, S., Khalid, K., Roslan, R. (2018). Modified Multi Prime RSA Cryptosystem. Journal of Physics: Conference Series, 995, 012030. https://doi.org/10.1088/1742-6596/995/1/012030
21. Farooq, S., Altaf, A., Iqbal, F., Thompson, E. B., Vargas, D. L. R., Díez, I. de la T., Ashraf, I. (2023). Resilience Optimization of Post-Quantum Cryptography Key Encapsulation Algorithms. Sensors, 23 (12), 5379. https://doi.org/10.3390/s23125379
22. Libert, B., Ling, S., Nguyen, K., Wang, H. (2023). Zero-Knowledge Arguments for Lattice-Based Accumulators: Logarithmic-Size Ring Signatures and Group Signatures Without Trapdoors. Journal of Cryptology, 36 (3). https://doi.org/10.1007/s00145-023-09470-6
23. Liang, W., Liu, Z., Zhao, X., Yang, Y., Liang, Z. (2024). Flexible and Compact MLWE-Based KEM. Mathematics, 12 (11), 1769. https://doi.org/10.3390/math12111769
24. Kieu-Do-Nguyen, B., The Binh, N., Pham-Quoc, C., Nghi, H. P., Tran, N.-T., Hoang, T.-T., Pham, C.-K. (2024). Compact and Low-Latency FPGA-Based Number Theoretic Transform Architecture for CRYSTALS Kyber Postquantum Cryptography Scheme. Information, 15 (7), 400. https://doi.org/10.3390/info15070400