IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 14, 2020, accepted May 24, 2020, date of publication June 5, 2020, date of current version June 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000316

Efficient NewHope Cryptography Based Facial Security System on a GPU

PHAP DUONG-NGOC, TUY NGUYEN TAN, (Member, IEEE), AND HANHO LEE, (Senior Member, IEEE)
Department of Information and Communication Engineering, Inha University, Incheon 22212, South Korea

Corresponding author: Hanho Lee (hhlee@ inha.ac)

This work was supported by the Basic Science Research Program funded by the Ministry of Science, ICT, South Korea, through the National Research Fund under Grant 2019R1F1A1061926.

ABSTRACT With explosive era of machine learning development, human data, such as biometric images, videos, and particularly facial information, have become an essential training resource. The popularity of video surveillance systems and growing use of facial images have increased the risk of leaking personal information. On the other hand, traditional cryptography systems are still expensive, time consuming, and low security, leading to be threatened by the foreseeable attacks of quantum computers. This paper proposes a novel approach to fully protect facial images extracted from videos based on a post-quantum cryptosystem named NewHope cryptography. Applying the proposed technique to arrange input data for encryption and decryption processes significantly reduces encryption and decryption times. The proposed facial security system was successfully accelerated using data-parallel computing model on the recently launched Nvidia GTX 2080Ti Graphics Processing Unit (GPU). Average face frame extracted from video (190 × 190 pixel) required only 2.2 *ms* and 2.7 *ms* total encryption and decryption times with security parameters $n = 1024$ and $n = 2048$, respectively, which is approximately 9 times faster than previous approaches. Analysis results of security criteria proved that the proposed system offered comparable confidentiality to previous systems.

INDEX TERMS Cryptosystem, facial security system, graphics processing unit, NewHope, public-key encryption.

I. INTRODUCTION

The digital age has penetrated people's lives and created an information explosion with tremendously increasing multimedia data. This has motivated machine learning algorithms development to best address big data problems [1]. However, these algorithms often require to run on remote cloud servers with powerful computing capabilities and collect private user information for the predictive models. This leads to growing concerns regarding user data privacy as well as numerous opportunities for attackers and other malicious actors to compromise sensitive information. These challenges have motivated researchers to develop a cryptography based cloud computing system that supports encrypted data to protect privacy sensitive user information, e.g., facial images, ensuring user information is encrypted before being sent to cloud based servers, as shown in Fig. 1.

Shor [2] showed that current cryptographic algorithms can be solved in constant time with quantum computers. Therefore, post-quantum algorithms have become widespread and are being considered as promising solutions to deal with potential attacks and provide provable security to safeguard sensitive data from the quantum computer threats. Lattice-based cryptography [3] has appeared as a modern security foundation. Its constructions dominated for the second round of the National Institute of Standards and Technology (NIST) post-quantum cryptography (PQC) standardization process, comprising promising candidates for PQC with strong theoretical security guarantees. Lattice constructions also provide initial primitives for many cryptographic functionalities, e.g., fully homomorphic encryption [4]. With the same key size, lattice-based cryptographic algorithms can offer much higher secure compared with conventional public-key algorithms, such as the Rivest–Shamir–Adleman (RSA) cryptosystem [5], or Elliptic Curve Cryptography (ECC) algorithms [6].

Most of the practical lattice-based cryptosystems are built upon ring learning with errors (R-LWE) problem. Several software and hardware implementations of R-LWE cryptography and biometric cryptosystems have been

The associate editor coordinating the review of this manuscript and approving it for publication was Chien-Ming Chen.

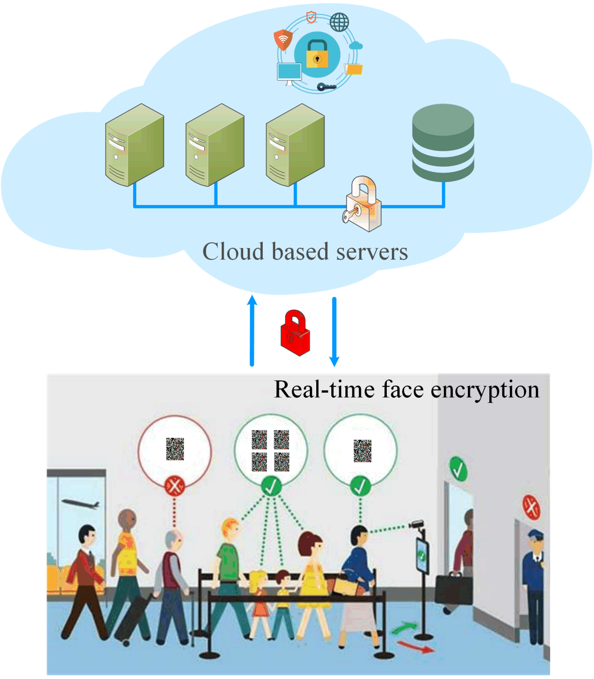


FIGURE 1. **Post-quantum cryptography based facial security system.**

reported [7]–[14]. In R-LWE primitives, polynomial multiplication is the basic and most expensive computation, and the number theoretic transform (NTT) is often utilized to perform polynomial multiplication efficiently.

Roy *et al.* [7] proposed a general hardware architecture for R-LWE based cryptography compared with ECC and N-th degree Truncated polynomial Ring Units (NTRU) cryptosystems. This approach decreased the complexity of forward NTT by avoiding the inherent pre-computed step in negative wrapped convolution (NWC) technique. However, this optimization cannot be applied to inverse NTT (INTT) due to the selection of transformation algorithms. Thomas *et al.* [8] proposed a software implementation on an 8-bit processor for the signature scheme using improved NTT approach. This method further removed the post-processing step of INTT and hiding the scaling of n^{-1} with a pre-calculated NTT, but not eliminated. Applying the R-LWE scheme for biomedical cryptosystems, Tan and Lee [9], [10] proposed the authentication systems based on biometrics, i.e., fingerprints, iris, and palmprints. These approaches encrypted biomedical images before sending to authentication systems, enhancing confidentiality over networks. Unfortunately, slow processing speed and low security may limit their scheme for practical applications. Tan *et al.* [11] also proposed a face encryption and decryption system effectively accelerated on a GPU platform but did not clearly indicate to the system security criteria.

Additionally, Liu *et al.* [12] proposed an efficient method based on spatial and value scrambling models to preserve face regions on surveillance videos. This scheme was able to localize and encrypt multiple faces well with satisfied anti-attack property. Thiyagarajan *et al.* [13] used the logistic map and a complex diffusion matrix in a chaotic image encryption scheme. Their encryption methodology offered a high resistance to statistical and differential attacks. Luo *et al.* [14] presented an efficient image secure solution based on ECC and public key encryption, namely EC-ElGamal, and chaotic theory to improve the randomness of the pixel distribution in cipher images. However, these current cryptosystems are still cumbersome and time consuming, which transit toward PQC to make image cryptosystems feasible and practical.

At Usenix 2016, Alkim *et al.* [15] proposed the NewHope post-quantum key agreement scheme, which was subsequently proposed to NIST PQC for standardization. NewHope is a lattice-based candidate for key-establishment in the second round of the NIST PQC standardization process [16]. NewHope's security is based on the conjectured quantum hardness of the R-LWE problem, and currently no attacks are known that can exploit the addition structures. Benchmark results confirmed that NewHope was computationally inexpensive with only a slight latency increases for some slow internet connections. This study leverages NewHope strong points as the basis for a high-speed video-based facial cryptosystem.

Parallel programming is being widely adopted for various computer research and development areas to improve algorithm performance while ensuring optimal results. Heterogeneous computing developments have facilitated adopting graphics processing units (GPUs) for parallel processing applications, taking advantage of their multithread-multicore architecture providing tremendous computation power and high memory bandwidth to effectively deal with scalable multi-objective decision variable problems. NVIDIA corp. introduced compute unified device architecture (CUDA) based on C/C++ in 2006 and the latest guide [17] helps to develop high speed applications that transparently scale parallelism to leverage the large number of processor cores on modern GPU devices. Parallel computing has been utilized to implement real-time video processing with transparent integration and optimization.

In this paper, we propose a NewHope cryptography based highly secure low latency facial security system to guard against adversarial attacks. Practical implementation confirmed that the proposed approach outperformed conventional works in term of processing time and had comparable information confidentiality. The main contributions of this paper are as follows:

1) The proposed approach offers a high efficiency solution to fully protect color facial images extracted from real-time video streams. The model provides high accuracy and privacy protection based on NewHope cryptography.
2) With major encoding and decoding refinements, the proposed scheme pushes the security level up and shows how to dramatically accelerate the encryption

and decryption processes, hence significantly reduce overall cryptosystem processing time.

3) This proposed scheme fully exploits various GPU platform's parallel architecture, and benchmark results confirm the scheme's applicability for a wide range of real-time image encrypting applications.
4) The resulting encrypted data completely satisfies security theory criteria, including similarity, information entropy, correlation coefficient (CC), number of pixels change rate (NPCR), and unified average changing intensity (UACI).

The remainder of this paper is constructed as follows: Section II briefly introduces the fundamental of R-LWE based public key cryptography, NTT multiplier, and random noise generator basic considerations. Section III describes the proposed NewHope cryptography based facial security system, and Section IV presents implementation results and performance analyses. Finally, conclusions are given in Section V.

II. BACKGROUND

A. R-LWE BASED CRYPTOGRAPHY

Regev [18] proposed the Learning with Errors (LWE) problem and showed that it is difficult to solve worst-cases under quantum reduction. However, these LWE-based practical applications are inefficient due to inherent complex mathematics. To achieve efficient computation and reduce key size, Lyubashevsky *et al.* [19] proposed R-LWE, an ideal lattice algebraic LWE variant utilizing special structures, which helps to withstand quantum computer based attacks.

For the R-LWE problem, we define polynomials in ring $Z[x]/<f(x)>$ where $f(x)$ is an irreducible polynomial of chosen degree n. For efficiency, define $f(x) = x^n + 1$ and let $R_q = Z_q/(x^n + 1)$ be the ring of integer polynomials modulo polynomial $(x^n + 1)$ with coefficients reduced modulo q. The general scheme based on the R-LWE problem [7]–[10], [19] includes three major procedures as follows:

1) Key generation: Sample polynomials $a, e, s \in R_q$ and compute $b = a \times s + e \in R_q$, where (a, b) and s are the public-key and the private key, respectively.
2) Encryption: Input message m is first encoded to $m_e \in R_q$, and then polynomials $e_1, e_2, e_3 \in R_q$ are randomly sampled. Ciphertext can be obtained as $c_1 = a \times e_1 + e_2$ and $c_2 = p \times e_1 + e_3 + m_e \in R_q$.
3) Decryption: Compute $m_d = c_1 \times s + c_2 \in R_q$ and recover the original message m from m_d using a decoder.

However, the most critical obstacle of R-LWE encryption still persist, i.e., large ciphertext expansion, resulting in huge amounts of calculations and transformations. Newhope [15] and HILA [20] offer many effective modifications for polynomial multiplication and ciphertext data compression to reduce R-LWE scheme arithmetic complexity for public key encryption problems. Processing input pixels can also take considerable computational overhead [9]–[11], particularly for high resolution images. This paper exploited an efficient input data arrangement with chosen polynomial degree for deployment on different platforms.

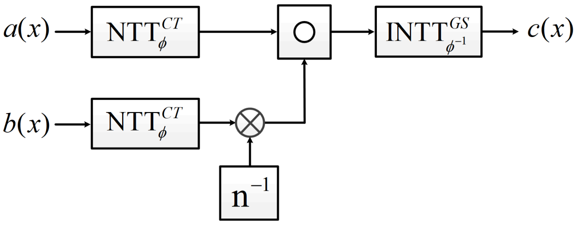

$a(x)$
NTT_{ϕ}^{CT}
$\circ$
$\mathrm{INTT}_{\phi^{-1}}^{GS}$
$c(x)$
$b(x)$
NTT_{ϕ}^{CT}
$\otimes$
n^{-1}

FIGURE 2. Number theoretic transform multiplier [8].

B. NUMBER THEORETIC TRANSFORM MULTIPLIER

The NTT [21] is a form of discrete fast Fourier transformation, providing a strong tool for polynomial multiplication in a finite field with integers rather than complex numbers.

The original NTT-based multiplier in R_q domain needs zero padding which is doubling a and b input sizes. We utilize the NWC technique [8], [10], [22] to avoid zero padding. We choose n as a power of two and define a primitive $2n$-th root of unity $\phi \in Z_q$, i.e., $\phi^2 = \omega \bmod q$ where q is a modulus prime in form $q = 1 \bmod 2n$. The NTT can be implemented efficiently by merging multiplications of powers of ω with powers of ϕ and ϕ^{-1}, respectively, to eliminate pre-computed and post-processing steps. The NWC theorem reduces the degree of polynomials in NTT and INTT from $2n$ to n and eliminates modulo $(x^n + 1)$.

Using Cooley-Turkey (CT) [23] and Gentleman-Sande (GS) [24] algorithms to compute NTT and INTT, respectively, can help avoid expensive bit-reverse steps when performing polynomial multiplication, $c = a \times b$, in R_q domain as shown in Fig. 2 with details:

$$c = INTT_{\phi^{-1}}^{GS}(NTT_{\phi}^{CT}(a) \circ NTT_{\phi}^{CT}(b)), \tag{1}$$

where the symbol $\circ$ denotes coefficient-wise multiplication. For a polynomial $a = (a[0], \ldots, a[n-1])$ we define:

$$NTT_{\phi}^{CT}(a) = \hat{a}[i] = \phi^i \sum_{j=0}^{n-1} a[j]\omega^{ij} \bmod q, \tag{2}$$

$$INTT_{\phi^{-1}}^{GS}(\hat{a}) = a[i] = n^{-1}\phi^{-i} \sum_{j=0}^{n-1} \hat{a}[j]\omega^{-ij} \bmod q. \tag{3}$$

C. RANDOM NOISE GENERATOR

The R-LWE problem hardness is directly related to the sampled polynomials' statistical properties. This provides an accurate and efficient sampler, which is a critical component for any lattice cryptographic implementation. Several methods are exploited for R-LWE cryptosystem operations to conduct random polynomial sampling. Although discrete Gaussian rejection and inversion sampling are popular choices, there are significant challenges to implement

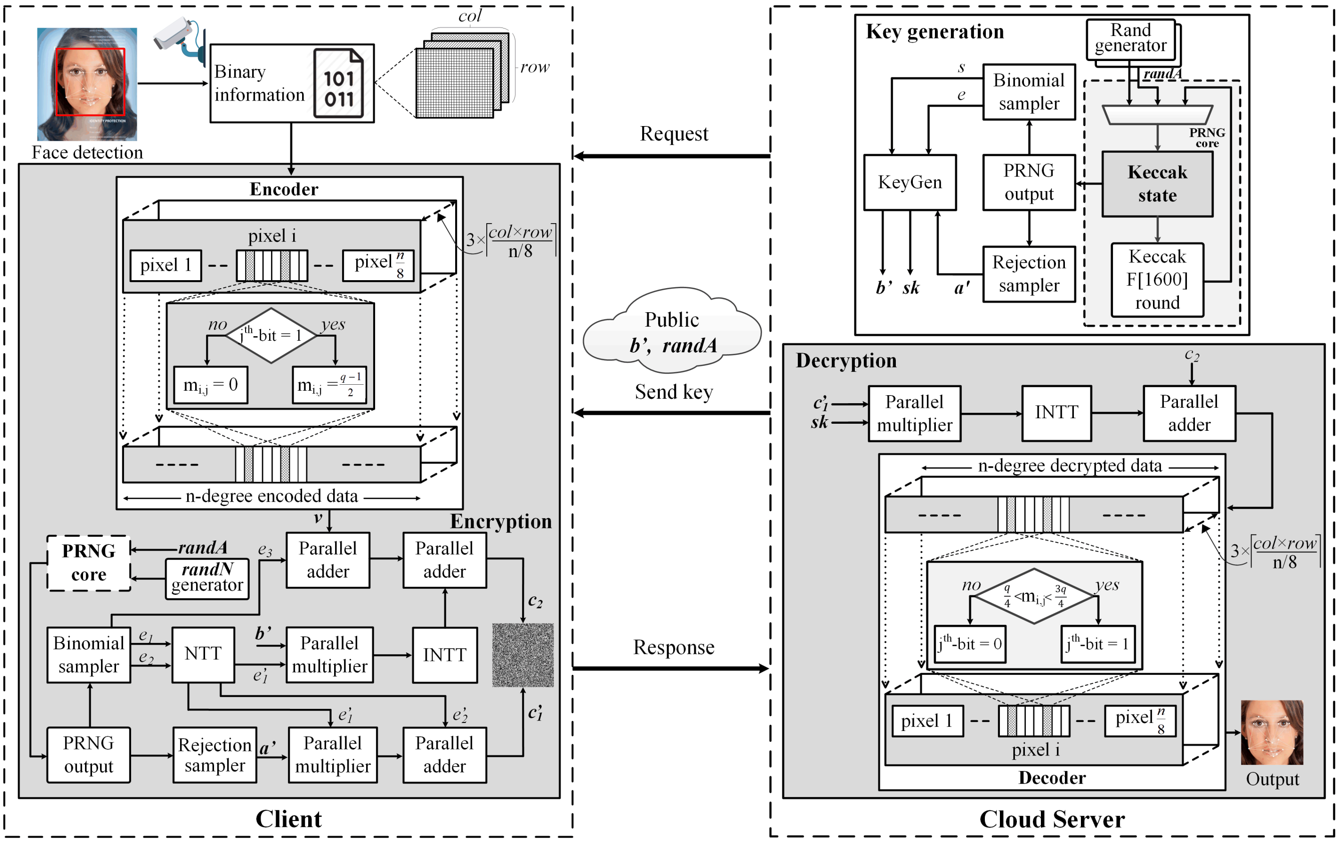


FIGURE 3. Proposed NewHope cryptography based facial security system.

```
Algorithm 1 Key Generation (Server)
Precondition: 32-byte random array
Output: public key ($pk$) and secrete key ($sk$)
$randA \xleftarrow{\$} \{0, \ldots, 255\}^{32}$
$a' \xleftarrow{\$} f^n(randA)$
$s, e \xleftarrow{\$} \Psi_k^n(\{0, 1\}^{32})$
$sk \longleftarrow NTT(s)$
$e' \longleftarrow NTT(e)$
$b' : = a' \circ sk + e'$
$pk : = (b', randA)$
return $(pk, sk)$
```

a discrete Gaussian sampler effectively and protect against timing side-channel attacks [15], [25], [26]. The high rejection rate of samples along with the expensive calculation of *exp*() are the main reasons for inefficiency. Bernoulli sampler introduced in [27] is an optimized version of rejection operation with no need to calculate the *exp*() function or pre-computed probabilities. However, the time dependency of Bernoulli makes it vulnerable to timing attacks. The Ziggurat sampler is a variation of the rejection sampler introduced in [28] for a continuous Gaussian sampler, which offers a certain tradeoff between memory consumption and performance. More precision imposes heavy memory overhead, which could drop performance if the cache is fully occupied. Knuth-Yao sampler [29] provides a near optimal sampling due to near entropy consumption of the random bits. However, the output of Knuth-Yao is generated within an unpredictable amount of time [30]. Interestingly, Göttert *et al.* [31] first employed the rejection sampler to generate uniform random parameters for lattice-based cryptography, which can achieve remarkable speed and area improvement using lazy floating-point arithmetic. This rejection sampler was subsequently used in NewHope [15], and further optimized by Gueron and Schlieker [32] to provide 1.5 times performance improvement. Gueron *et al.* improved the sampling operation using the full 16-bit sample and reducing the rejection rate from 25% to 6%. This proposal develops a fast sampler using Keccak permutation based extendable-output function (XOF) [33] with a small probability of rejection. Public parameters for R-LWE schemes are commonly generated by one party and sent to the other through a transmission channel, which consumes considerable network bandwidth. More effective communication would be to let both parties use the same sampler to generate the public parameters independently from a shared random seed.

Additionally, NewHope manipulates the centered binomial distribution ψ_k for the secret and noise term, which eliminate the expensive calculation of *exp*() and pre-computed table in the rounded Gaussian sampler. The ψ_k is computed by $\sum_{i=0}^{k-1}(b_i - b_i')$, where b_i and $b_i' \in \{0, 1\}$ are uniform independent bits, and $k = 8$ for all instantiations [15]. Sampling from the binomial distribution ψ_k involves $2k$ random bits take values in range of $[-k, k]$. The distribution ψ_k is zero centered, with variance k and standard deviation $\sigma = \sqrt{k/2}$. The choice of variance k can make a tradeoff between error probability and security level. Choosing a higher variance for binomial sampler, proposed scheme leaves a comfortable margin to the targeted 256 post-quantum bit-security. Binomial sampler has been also employed for many software implementations of HILA5 [20], LAC [34], LIMA [35], Kyber [36], and Titanium [37].

III. PROPOSED CRYPTOSYSTEM FOR VIDEO-BASED FACIAL SECURITY

A. CRYPTOGRAPHY SCHEME FOR FACIAL SECURITY

Fig. 3 describes the proposed efficient cryptosystem for video based facial security. The most outstanding features are efficient encoding and decoding, and parallel processing for multiple faces frame encryption and decryption. Ciphertext size can be significantly reduced by employing an effective coefficient arrangement for transferring over communication channels. Moreover, public parameter properties and proper NTT operation placement differ from previous approaches [7]–[11], saving four NTT and two INTT computations.

Faces appearing in the video were extracted using Haar-based high speed support functions from the OpenCV library [38], [39] by the client. Face regions were then input into the encryption process before sending to a remote server over the network. These cipher images were then decrypted by the cloud server for specific applications as required.

1) KEY GENERATION

Algorithm 1 shows the key generation procedure, where the server uses an array of 32 random bytes as input for the SHAKE256 XOF sampling function [33] to generate a high security level public parameter a (256-bit security) with uniform distribution coefficients. We randomly generate $a \in [0, q)$ by uniform sampler in GF(q), and it is usually freshly created for each instance to resist against various attack modes. With uniform distribution in GF(q), a can be considered as directly sampled in the NTT form. Hence, two NTT steps of polynomial a can be omitted in the key generation procedure.

Additionally, we use a 32-bit random string input for the binomial sampling function. These random noise polynomials are sampled over a centered binomial error distribution. We then calculate the public parameter b' in NTT format, and combine it with a' in the public key sent to the client. Rather than transmit the complete polynomial a', the server requires only its random seed, hence reducing the total public-key size.

Algorithm 2 Facial Image Encryption (Client)

Input: public key (pk) and input image (m)
Output: ciphertext c_1' and c_2

```
$a' \xleftarrow{\$} f^n(randA)$
$e_1, e_2, e_3 \xleftarrow{\$} \Psi_k^n(\{0,1\}^{32})$
$e_1' \longleftarrow ParallelNTT(e_1)$
$e_2' \longleftarrow ParallelNTT(e_2)$
$m = Concat(image)$
$c \longleftarrow [0 \div 3 \times \lceil \frac{col \times row}{n/8} \rceil - 1]$
$p \longleftarrow [c \times \frac{n}{8} \div (c+1) \times \frac{n}{8} - 1]$
//ParallelEncoder($m$)
for $i = 0$ to $n-1$ do
    for $j = 0$ to $7$ do
        $m_e[i] : = \lfloor q/2 \rfloor \times m[8 \times p + j]$
    end for
end for
$c_1' : = a' \circ e_1' + e_2'$
$c_2 : = ParallelINTT(b' \circ e_1') + e_3 + m_e$
return $(c_1', c_2)$
```

2) ENCRYPTION

Algorithm 2 shows the face image encryption process at the client side. After receiving the random array from the server, the client pseudo-random noise generator (PRNG) component first generates the public parameter a using the same function as the remote server. Noise polynomials are also sampled using the binomial sampling function with 32-bit random string.

Encoding: For each input frame, three facial region channels were concatenated into an $3 \times (col \times row)$ pixel array using *Concat*, which is then separated into $n/8$-pixel groups and mapped sequentially for encoding. Depending on the given polynomial degree, the number of groups are calculated by $3 \times \lceil \frac{col \times row}{n/8} \rceil$. The round-up operation is with zero padding to retain pixels at the back of each channel. These mapping operations can be simultaneously applied on parallel architectures to reduce the encoding time as shown in Fig. 3 in the Encoder block. The final encoding step is to assign $m[i] \times (q-1)/2$ to each i^{th}-bit of $n/8$-pixel message m for corresponding polynomial coefficient.

The main concepts for this procedure include adding the n-bit encoded input message m_e in free NTT notation to generate the second ciphertext c_2. The first ciphertext c_1' is calculated by $e_2' + a' \circ e_1'$ in NTT form. Polynomials c_2 generated from n-bit input polynomials would be sent to server together with c_1' as ciphertexts of the original face images.

3) DECRYPTION

Algorithm 3 shows the server decryption process for ciphertext pairs (c_1', c_2) received from clients and the secret key generated in Algorithm 1. Decryption is only possible with

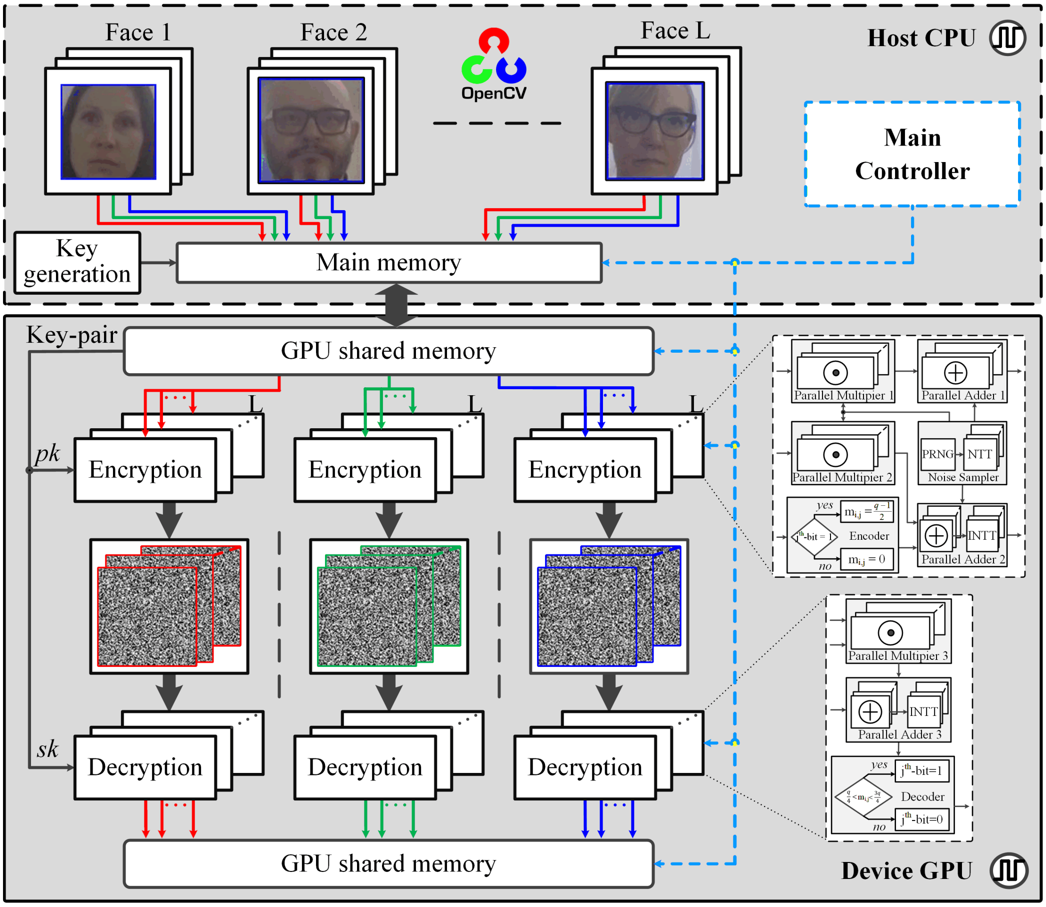


FIGURE 4. Proposed parallel architecture for NewHope cryptography based facial security system on a GPU platform.

knowledge of sk, since the large term b' cannot be eliminated when computing $c_2 - INTT(c_1' \circ sk)$. Thus, this computation hides the original binary message in decryption.

Decoding: Decoding performs the opposite steps to encoding to reconstruct the original image. Using the chosen encode threshold from encryption side, the decoding function checks whether each received coefficient $m_d[i]$ is in the interval $\lfloor q/4, 3 \times q/4 \rfloor$, which is interpreted as 1 or 0. Decryption data obtained from the Algorithm 3 are completely identical to the original images.

B. PARAMETER SELECTION

Decryption error probabilities depend on the random noises e_1, e_2, and e_3 standard deviation σ. The noise follows a binomial random distribution ψ_k with carefully chosen variance, e.g., noise distribution ψ_k with variance $k/2$, has $\sigma = \sqrt{k/2}$. Smaller σ means smaller error probability, but also reduces security. Therefore, to obtain a fair decryption, we chose $k = 16$ [20].

The proposed scheme security is based on the apparent R-LWE problem hardness. It has been proven that the R-LWE problem is as hard as solving worst case lattice problems for certain parameters. The final scheme security depends on the set (n, q, k), where n is the polynomial degree, q is the coefficient modulus, and k is the integer parameter for the noise distribution. Larger n and larger k/q increase security. On the other hand, increasing q increases the noise budget, but lowers security, increases key size, ciphertext expansion, and reduces performance. To realize NTT efficiently, we chose the smallest prime modulus (14 bits). For the above q and k parameters, we recommend $n = 1024$ and higher for long term security. The proposed scheme can achieve 256-bit security level for that tuple set [15].

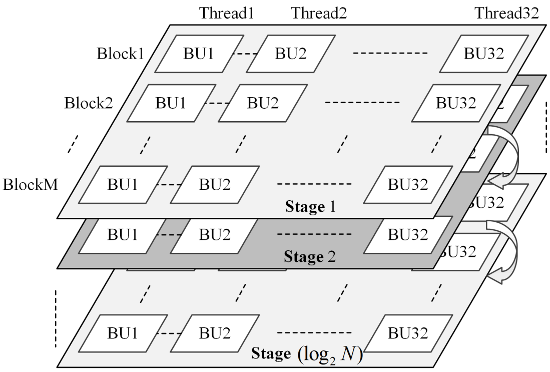
The chosen q means that each ring coefficient fits into 14 bits. To efficiently reduce data size for transmission or storage, we specify a method of packing a vector with n 14-bit coefficients into $14 \times n/8$ bytes. Each 14-bit segment can be concatenated into a continuous byte sequence in little-endian fashion, where the easiest encoding is blocks of eight coefficients. This helps to reduce public-key and ciphertext sizes for transmission between clients and servers by approximately 12.5%.

C. PARALLEL ARCHITECTURE OF SCHEME ON A GPU

Processing a high-resolution image often takes a long time and requires an efficient mechanism for any practical application. To address this problem, we harness the thousands of cores present inside a GPU. GPU cores can

Algorithm 3 Facial Image Decryption (Server)

```
Input: ciphertext $(c'_1, c_2)$ and secret key $sk$
Output: reconstructed image $(m')$
$m_d :\, = c_2 - ParallelINTT(c'_1 \circ sk)$
$c \longleftarrow [0 \div 3 \times \lceil \frac{col \times row}{n/8} \rceil - 1]$
$p \longleftarrow [c \times \frac{n}{8} \div (c+1) \times \frac{n}{8} - 1]$
//ParallelDecoder($m_d$)
for $i = 0$ to $n - 1$ do
    for $j = 0$ to $7$ do
        if $m_d[i] \in \lfloor \frac{q}{4} \div 3 \times \frac{q}{4} \rfloor$ then
            $m'[8 \times p + j] :\, = 1$
        else
            $m'[8 \times p + j] :\, = 0$
        end if
    end for
end for
return $(m')$
```



*BU: Butterfly Unit

FIGURE 5. Proposed parallel NTT operation on a GPU platform.

act independently and process different parts of the image in parallel. Fig. 4 shows the proposed parallel scheme for multiple face regions. NTT-based multiplications often take considerable processing time in R-LWE based schemes due to the sequence of butterfly units in different processing stages. Butterfly unit properties are independent in the same stage as shown in Fig. 5. Hence we can spread the total butterfly units in each stage running in parallel with different weights. Taking 1024-point polynomial for example, 16 blocks (M) with 32-thread each were run in parallel for each stage. This method can accelerate both NTT and INTT operations. We further merged the scaling of n^{-1} into the last round of INTT operation to eliminate expensive modular multiplication. We used radix-2 NTT multipliers, hence $N/2$ butterfly units may be processed simultaneously in each stage, which helps approximately N times speed up compared to continuity processing. Coefficient-wise addition and multiplication are also performed in parallel according to the polynomial degree.

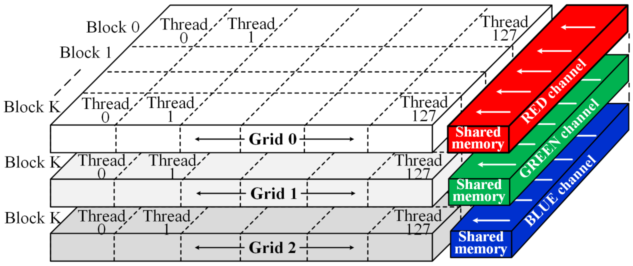


FIGURE 6. Proposed scheme to process color image in parallel on a GPU platform.

Additionally, encryption and decryption operations for color images were separated into $3 \times \lceil \frac{col \times row}{n/8} \rceil$ parts and run in parallel along with shared memory, as shown in Fig. 6. This parallel architecture can also help achieve real-time multiple faces encryption speeds for high definition videos, which would have significant implications in increasing processing speed for actual machine learning systems.

IV. PERFORMANCE EVALUATION AND COMPARISON

A. SIMULATION ENVIRONMENT

To evaluate the proposed cryptosystem effectiveness, all implementations were on a computer system comprising CPU Intel i9-999KF, 32 GB DRAM, Windows 10 64-bit operating system, and NVIDIA GeForce RTX 2080Ti GPU with 8 GB DRAM. We used CUDA Toolkit v10.1.243 and OpenCV v4.1.2. The proposed scheme implementation was evaluated on GPU and CPU platforms for images and videos to compare with previous studies.

B. TIME COMPLEXITY ANALYSIS

We evaluated the proposed scheme effectiveness for encryption and decryption using detected face regions from a video [40]. The face detection algorithm was run on an approximately 2 minutes long input video with 720 × 432 pixel resolution in MPEG-4 format at 30 frames per second.

Table 1 compares the proposed scheme processing time with previous schemes. Face regions were extracted from video with average size 190 × 190 pixel, and simulation results confirmed encryption and decryption required only 1.5 *ms* and 0.7 *ms* for $n = 1024$, and 1.8 *ms* and 0.9 *ms* for $n = 2048$, respectively, on the GPU platform. This outperformed Tan *et al.* [11] by approximately 90% and 87% for $n = 1024$ and $n = 2048$, respectively, in term of total processing time, while offering higher security. The evaluation results on CPU showed that our scheme was 6×, 14×, 13× and 5×, 12×, 11× faster than previous studies [9], [11], [14] for $n = 1024$ and $n = 2048$, respectively. Implementations were also deployed on various GPU platforms with different input frame compression rates into 2048 bit messages, as shown in Fig. 7. The proposed scheme achieved very low processing time on specific GPU Geforce RTX 2080Ti (2.7 *ms*) for the highest compression rate.

TABLE 1. Proposed scheme execution times for face encryption and decryption.

Parameters	CPU implementation					GPU implementation		
	Ref. [9]	Ref. [11]	Ref. [14]	This work		Ref. [11] (GTX TiTan Black)	This work (RTX 2080Ti)	
Domain	$n = 8$	$n = 256$	ECC	$n = 1024$	$n = 2048$	$n = 256$	$n = 1024$	$n = 2048$
Random sampler	Gaussian	Gaussian	SHA-2	SHA-3	SHA-3	Gaussian	SHA-3	SHA-3
Encryption (ms)	858	1500	1330	95	112	18	1.5	1.8
Decryption (ms)	62	700	626	58	64	2	0.7	0.9
Total (ms)	920	2200	1956*	153	179	20	2.2	2.7

* Normalized processing time for 3-channel frame size of 190×190 pixel

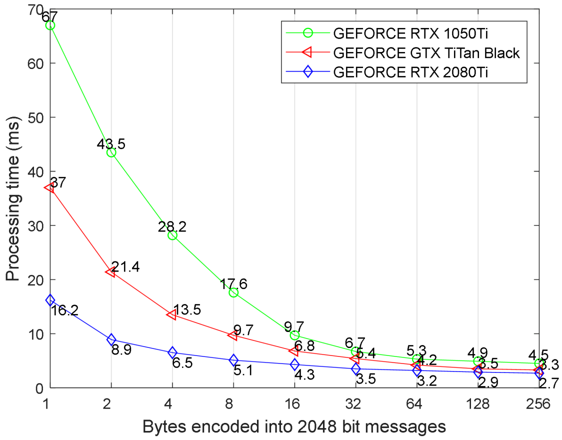


FIGURE 7. Comparison in execution times of face encryption and decryption on different GPUs with different encoded bytes for frame size of 190 × 190 pixel.

Fig. 8 shows applying the proposed system for digital video content protection. For frames contain multiple faces, face regions are detected and encrypted when each frame is input using the location of identify regions. The proposed approach encrypted three-face images simultaneously, requiring only 3.5 *ms* and 3.8 *ms* total processing time with $n = 1024$ and $n = 2048$, respectively. For content decryption, we extract the location information and embed the decrypted contents of digital video. The proposed system achieved real-time face region encryption and decryption with high performance, adapted for practical applications.

C. SECURITY LEVEL ANALYSIS

In addition to the histogram analysis was carried out on facial color images, we evaluated proposed cryptosystem on *Lena* gray scales image with size of 512 × 512 pixel to compare the effectiveness and security with previous studies. We employed several security measures based on different criteria to verify the proposed system robustness, including information entropy, CC, NPCR, and UACI.

1) HISTOGRAM ANALYSIS

In image processing contexts, histogram represents the pixel intensity distribution, showing the number of pixels corresponding to tonal values in a digital image. Pixel values range from 0 to 255 for 8-bit images, hence an efficient image encryption system must produce an apparently random-like cipher images with uniformly distributed pixel density in this range. Fig. 9 shows differences between detected face region histograms with the corresponding encrypted image. The homogeneous histograms confirm that the proposed scheme can resist statistical analysis attacks.

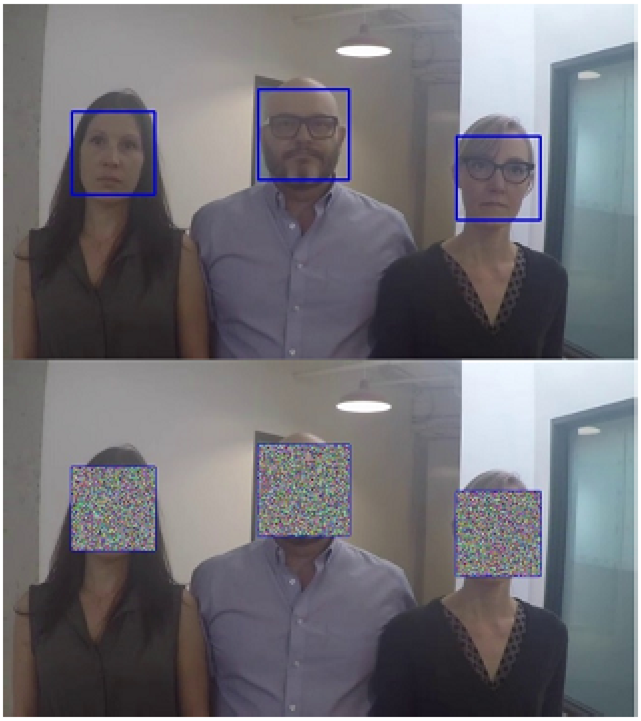


FIGURE 8. Encryption for multiple faces and digital video content protection.

2) CORRELATION COEFFICIENTS

The correlation results between two adjacent pixels in plain and cipher image calculated by:

$$CC_{xy} = \frac{cov(x, y)}{\sqrt{D(x)D(y)}}, \tag{4}$$

where

$$cov(x, y) = \frac{1}{N} \sum\nolimits_{i=1}^{N} (x_i - E(x_i))(y_i - E(y_i)),$$

TABLE 2. Comparison in correlation factor and information entropy.

Method	Correlation coefficients						Entropy (bits/pixel)	
	Diagonal		Vertical		Horizontal			
	Plain	Cipher	Plain	Cipher	Plain	Cipher	Plain	Cipher
Ref. [13]	–	0.0087	–	0.0011	–	0.0048	–	7.9943
Ref. [14]	0.9669	0.0011	0.9801	-0.0024	0.9858	0.0019	7.3871	7.9993
Proposed	0.9593	0.0039	0.9850	0.0015	0.9719	0.0002	7.3451	7.9978

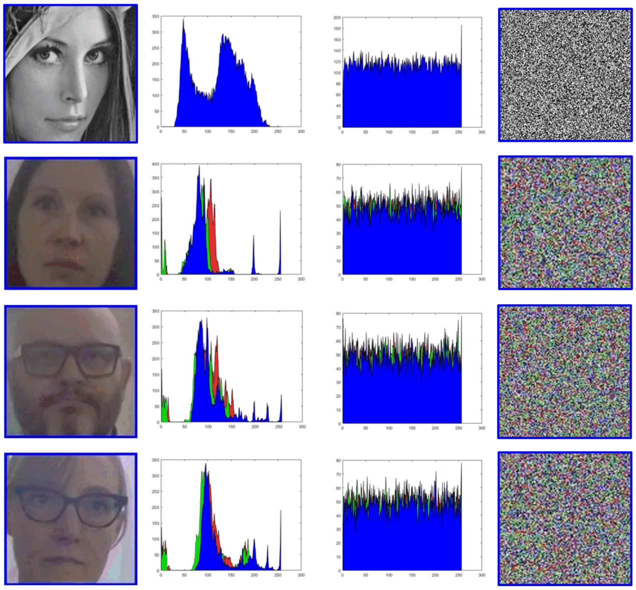


FIGURE 9. Typical original and encrypted facial images and corresponding histograms.

$$D(x) = \frac{1}{N}\sum\nolimits_{i=1}^{N}(x_i - E(x_i))^2,$$
$$E(x) = \frac{1}{N}\sum\nolimits_{i=1}^{N}x_i,$$

where x and y are 8-bit gray values for two adjacent pixels in an image; N is the total number of pixels; $D(x)$ and $E(x)$ are the variance and expectation of x, respectively.

Larger $|CC|$ in range of $[0, 1]$ implies better correlation between adjacent pixels in an image, e.g., stronger similarity. Adjacent pixels in the original image are usually highly correlated diagonally, vertically and horizontally, i.e., CC values are usually close to 1; whereas a good encryption algorithm should not retain any relationship between adjacent pixels in the encrypted image, i.e., CCs are expected to be close to 0.

Table 2 compares the *Lena* gray scale image CCs between proposed scheme to other studies. All cipher images CCs are close to 0, our proposed scheme has comparable results compared to [13], [14]. This result confirm that the proposed encryption can effectively reduce the correlation of adjacent pixels in cipher images.

3) INFORMATION ENTROPY ANALYSIS

Cipher images should also have high randomness to guarantee security. Information entropy is a useful metric to evaluate image randomness defined by:

$$H(s) = -\sum_{i=0}^{v-1} P(s_i)log_2(P(s_i)), \tag{5}$$

where s denotes the total number of symbols s_i, v is the number of values for each symbol and $P(s_i)$ is the occurrence probability for s_i; and $0 \leq H(s) \leq log_2 v$.

For an 8-bit gray image, higher $H(s)$ implies less predictable image values, with ideally $H(s)$ approaches 8. Table 2 shows that encrypted images achieve very high entropy, demonstrating the proposed scheme effectively hides the details of plain images. The comparison shows that the proposed method achieves higher value than [13] and slightly lower than [14], but the improvement in entropy from [14] (0.6122 bits/pixel) is less than that of ours (0.6526 bits/pixel). Hence the proposed scheme can produce cipher images with highly uniformly random distributions.

4) NPCR AND UACI ANALYSIS

Differential cryptanalysis is a powerful statistical attack takes advantage of non-uniform relations between plain image and corresponding cipher image. To determine a cipher's resistance to differential attacks, NPCR and UACI [41] are measured to evaluate the robustness of the proposed scheme. Taking two slightly different plain images, while NPCR calculates the change rate in number of pixels between two cipher images, UACI represents the mean variation of pixel strength in same location of two cipher images. The NPCR and UACI values are described as percentage and determined by:

$$NPCR = \frac{\sum_{i=1}^{W}\sum_{j=c1}^{H} D(i,j)}{W \times H} \times 100\%, \tag{6}$$
$$UACI = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} |C_1(i,j) - C_2(i,j)|}{F \times W \times H} \times 100\%, \tag{7}$$

where $C_1(i,j)$ and $C_2(i,j)$ denote pixels at same position of two cipher images, W and H are the width and height of these images, F is the maximal intensive value ($F = 255$ for gray

TABLE 3. Comparison in NPCR and UACI.

Method	NPCR (%)	UACI (%)
Ref. [13]*	99.6081	33.4321
Ref. [14]	99.6113	33.4682
Proposed	99.6104	33.4691

* Reported mean values 99.6073% and 33.4481% for NPCR and UACI

scale image), and $D(i, j)$ is change rate defined as:

$$D(i,j) = \begin{cases} 1, & if\ C_1(i,j) \neq C_2(i,j) \\ 0, & if\ C_1(i,j) = C_2(i,j). \end{cases}$$

In this test, we varied some random pixels in the original image, kept the same public key, generated freshly random noise, and evaluated our scheme on these plain images. NPCR and UACI values between two cipher images were calculated and furnished in Table 3. The expected values of NPCR and UACI are 99.6094% and 33.4635%, respectively, for [14] and our scheme. The mean values of [13] are reported slightly different to ours. Table 3 shows that NPCR and UACI values are very close to the mean for given image size. NPCR value of proposed scheme is larger than [13] and slightly smaller than [14] while our UACI is largest. These comparison results prove that our scheme is strength against differential attacks.

V. CONCLUSION

In this paper, we proposed an efficient NewHope cryptography based scheme for video-based facial security system, particularly suitable for a wide range of real-time vision applications. The main contributions from the proposed cryptosystem include high speed encoding and decoding techniques based on effectively arranging the pixel, and combine with NewHope scheme improvements for facial security. We also concentrated on accelerating the NTT-based multiplication and color images processing, taking full advantage of parallel architecture on GPU platform. The proposed system provides low latency and high security, offering a potential mechanism for real-time processing applications. Future works will explore the flexible parallel architecture of configurable devices, such as field-programmable gate array, for a higher performance image encryption system.

REFERENCES

[1] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, pp. 1–21, Feb. 2015.

[2] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM J. Comput.*, vol. 26, no. 5, pp. 1484–1509, Oct. 1997.

[3] I. Verbauwhede, J. Balasch, S. S. Roy, and A. Van Herrewege, "24.1 circuit challenges from cryptography," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 428–429.

[4] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proc. 41st Annu. ACM Symp. Symp. Theory Comput. (STOC)*, 2009, pp. 169–178.

[5] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public- key cryptosystems," *Wireless Pers. Commun.*, vol. 77, no. 2, pp. 907–922, Feb. 1978.

[6] T. T. Nguyen and H. Lee, "Efficient algorithm and architecture for elliptic curve cryptographic processor," *J. Semicond. Technol. Sci.*, vol. 16, no. 1, pp. 118–125, Feb. 2016.

[7] S. S. Roy, F. Vercauteren, N. Mentens, D. D. Chen, and I. Verbauwhede, "Compact ring-LWE Cryptoprocessor," in *Proc. CHES*, vol. 8731, Sep. 2014, pp. 371–391.

[8] P. Thomas, T. Oder, and G. Tim, "High-performance ideal lattice-based cryptography on 8-bit ATxmega microcontrollers," in *Proc. 4th Int. Conf. Cryptol. Inf. Secur. Latin Amer.*, Aug. 2015, vol. 9230, no. 645622, pp. 346–365.

[9] T. N. Tan and H. Lee, "High-performance ring-LWE cryptography scheme for biometric data security," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 2, pp. 97–106, Apr. 2018.

[10] T. N. Tan and H. Lee, "High-secure fingerprint authentication system using ring-LWE cryptography," *IEEE Access*, vol. 7, pp. 23379–23387, Feb. 2019.

[11] T. N. Tan, Y. Hyun, J. Kim, D. Choi, and H. Lee, "Ring-LWE based face encryption and decryption system on a GPU," in *Proc. Int. SoC Design Conf. (ISOCC)*, Oct. 2019, pp. 15–16.

[12] S. Liu, L. Kong, and H. Wang, "Face detection and encryption for privacy preserving in surveillance video," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, vol. 2, Nov. 2018, pp. 162–172.

[13] J. Thiyagarajan, B. Murugan, and A. Gounder, "A chaotic image encryption scheme with complex diffusion matrix for plain image sensitivity," *Serbian J. Electr. Eng.*, vol. 16, no. 2, pp. 247–265, 2019.

[14] Y. Luo, X. Ouyang, J.-X. Liu, and L.-C. Cao, "An image encryption method based on elliptic curve ElGamal encryption and chaotic systems," *IEEE Access*, vol. 7, pp. 38507–38522, 2019.

[15] E. Alkim, L. Ducas, T. Pöppelmann, and P. Schwabe, "Post-quantum key exchange—A new hope," in *Proc. 25th USENIX Secur. Symp.*, Aug. 2016, pp. 327–343.

[16] E. Alkim, R. Avanzi, J. Bos, L. Ducas, A. de la Piedra, T. Pöppelmann, P. Schwabe, and D. Stebila. (2019). *Newhope: Algorithm Specification and Supporting Documentation. Submission to the NIST Post-Quantum Cryptography Standardization Project*. [Online]. Available: https://csrc.nist.gov/projects/post-quantum-cryptography/round-2-submissions

[17] NVIDIA. (Nov. 2019).*Cuda C++ Programming Guide*. [Online]. Available: https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html

[18] O. Regev, "The learning with errors problem," *Proc. Annu. IEEE Conf. Comput. Complex.*, vol. 3, no. 015848, pp. 191–204, Jun. 2010.

[19] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 6110. Berlin, Germany: Springer, May 2010. pp. 1–23.

[20] D. J. Bernstein, L. G. Bruinderink, T. Lange, and L. Panny, "HILA5 pindakaas: On the CCA security of lattice-based encryption with error correction," in *Progress in Cryptology* (Lecture Notes in Computer Science), vol. 10831. Cham, Switzerland: Springer, May 2018, pp. 203–216.

[21] F. Winkler, *Polynomial Algorithms in Computer Algebra*. Vienna, Austria: Springer-Verlag, 1996, pp. 26–81.

[22] P. Longa and M. Naehrig, "Speeding up the number theoretic transform for faster ideal lattice-based cryptography," *Lect. Notes Comput. Sci.*, vol. 10052, pp. 124–139, Nov. 2016.

[23] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.

[24] W. M. Gentleman and G. Sande, "Fast Fourier Transforms: For fun and profit," in *Proc. AFIPS Conf. Proc. Fall Joint Comput. Conf.*, Nov. 1966, pp. 563–578.

[25] J. W. Bos, C. Costello, M. Naehrig, and D. Stebila, "Post-quantum key exchange for the TLS protocol from the ring learning with errors problem," in *Proc. IEEE Symp. Secur. Privacy*, May 2015, pp. 553–570.

[26] H. Nejatollahi, N. Dutt, S. Ray, F. Regazzoni, I. Banerjee, and R. Cammarota, "Post-quantum lattice-based cryptography implementations," *ACM Comput. Surveys*, vol. 51, no. 6, pp. 1–41, Feb. 2019.

[27] L. Ducas, A. Durmus, T. Lepoint, and V. Lyubashevsky, "Lattice signatures and bimodal Gaussians," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 8042. Berlin, Germany: Springer, Aug. 2013, pp. 40–56.

[28] G. Marsaglia and W. W. Tsang, "The ziggurat method for generating random variables," *J. Stat. Softw.*, vol. 5, no. 8, pp. 1–7, 2000.

[29] D. Knuth and A. Yao, "The complexity of nonuniform random number generation," in *Processing of Symposium on New Directions and Recent Results in Algorithms and Complexity*. New York, NY, USA: Academic, 1976, pp. 357–428.

[30] C. P. Renteria-Mejia and J. Velasco-Medina, "High-throughput ring-LWE cryptoprocessors," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 8, pp. 2332–2345, Aug. 2017.

[31] N. Göttert, T. Feller, M. Schneider, J. Buchmann, and S. Huss, "On the design of hardware building blocks for modern lattice-based encryption schemes," in *Proc. 14th Int. Workshop Cryptograph. Hardw. Embedded Syst. (CHES)*, Leuven, Belgium, vol. 7428, Sep. 2012, pp. 512–529.

[32] S. Gueron and F. Schlieker, "Speeding up R-LWE post-quantum key exchange," in *Proc. 21st Nordic Conf. Secure IT Syst.* (Lecture Notes in Computer Science), vol. 10014. Cham, Switzerland: Springer, 2016, pp. 187–198.

[33] M. J. Dworkin, "SHA-3 standard: Permutation-based hash and extendable-output functions," in *Proc. NIST*, Aug. 2015, pp. 7–24.

[34] X. Lu, Y. Liu, Z. Zhang, D. Jia, H. Xue, J. He, B. Li, and K. Wang, "LAC : Practical ring-LWE based public-key encryption with byte-level modulus," in *Proc. IACR*, Dec. 2018, p. 1009.

[35] N. P. Smart, M. R. Albrecht, and R. Holloway, "LIMA 1.1—A post quantum cryptography encryption scheme," in *Proc. Nat. Inst. Standards Technol.*, Jun. 2017, pp. 9–31.

[36] J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, P. Schwabe, G. Seiler, and D. Stehle, "CRYSTALS–kyber: A CCA-secure Module-Lattice-Based KEM," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroS&P)*, Apr. 2018, pp. 353–367.

[37] R. Steinfeld, A. Sakzad, and R. K. Zhao, "Titanium: Post-quantum public-key encryption and KEM algorithms," in *Proc. Nat. Inst. Standards Technol.*, May 2018, pp. 4–30.

[38] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Dec. 2001, p. 1.

[39] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, 2002, pp. 900–903.

[40] C. Contaoi. *Sample Videos*. Accessed: Aug. 24, 2019. [Online]. Available: https://github.com/intel-iot-devkit/sample-videos/blob/master/head-pose-face-detection-female-and-male.mp4

[41] G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos, Solitons Fractals*, vol. 21, no. 3, pp. 749–761, Jul. 2004.

PHAP DUONG-NGOC received the B.S. degree in electronics and telecommunication engineering from the Da Nang University of Technology, in 2009, and the M.S. degree in electronics engineering from Da Nang University, Vietnam, in 2015. He is currently pursuing the Ph.D. degree in information and communication engineering with Inha University. His research interests include algorithms and architecture design for cryptosystems and homomorphic encryption.

TUY NGUYEN TAN (Member, IEEE) received the M.S. and Ph.D. degrees in information and communication engineering from Inha University, South Korea, in 2016 and 2019, respectively. He was a Circuit Design Engineer at Silicon Design Solutions Company, in 2009, a Field Engineer at GTel Mobile JSC, from 2009 to 2013, and has been a Researcher with Ton Duc Thang University, Vietnam, since 2019. His research interests include VLSI algorithms and architecture design for cryptosystems and error correction codes.

HANHO LEE (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Minnesota, Minneapolis, in 1996 and 2000, respectively. He was a Member of Technical Staff at Lucent Technologies (Bell Labs Innovations), Allentown, from April 2000 to August 2002, and an Assistant Professor with the Department of Electrical and Computer Engineering, University of Connecticut, USA, from August 2002 to August 2004. He was a Visiting Scholar at Bell Labs and Alcatel-Lucent, NJ, USA, from August 2010 to August 2011. He has been with the Department of Information and Communication Engineering, Inha University, since August 2004, where he is currently a Professor. His research interests include algorithms and architecture design for cryptographic hardware, forward error correction coding, and digital signal processing.