HPU2 Journal of Sciences:
Natural Sciences and Technology

journal homepage: https://sj.hpu2.edu.vn

Article type: Research article

Privacy preserving using extended euclidean – algorithm applied to RSA

Thi-Nhung Nguyen*

University of Information and Communication Technology, Thai Nguyen University, Thai Nguyen, Vietnam

Abstract

RSA cryptography is a strong encryption method widely used in online transactions. Using the extended Euclidean algorithm is an important and efficient technique for finding the secret key in RSA cryptography. This study provides an implementation of the extended Euclidean algorithm to find secret keys based on RSA cryptography and hopes that it can be of help to experts in the field of information security.

Keywords: Cryptography, Steganography, Watermarking, RSA, private key, Public key.

1. Introduction

Nowadays, with the explosion of information technology, transmitting and encrypting information plays an important role in every human activity. It is a matter of life and death when information can be attacked and stolen. For Industrial revolution 4.0, there is an increasing need for information exchange between components in society. Furthermore, the need to connect with the outside world is also higher. As a result, the computer network was created, bringing many benefits to humans thanks to the quick and accurate exchange and process of information. Tasks can be solved promptly and efficiently anywhere on Earth. However, these conveniences also raise a question: Is the information sent from the sender to the recipient absolutely safe? Who can ensure that our information is not accessed illegally and is kept confidential? In the communication system, information stored, transmitted, and used on the public information network can be eavesdropped on, hijacked, distorted, or destroyed, leading to incalculable losses. Especially, the data from the banking system, commercial system, government management agencies or those in the military and diplomatic fields are exchanged

* Corresponding author, E-mail: ntnhung@ictu.edu.vn

https://doi.org/10.56764/hpu2.jos.2023.1.2.53-63

Received date: 05-4-2023 ; Revised date: 24-4-2023 ; Accepted date: 25-4-2023

in a secure and confidential manner. Therefore, data security is of paramount importance [2]. Cryptography plays an important role in providing security and ensuring safe data transmission over the Internet. One of the principles of cryptographic techniques is to provide security for sensitive information through digital signatures, authentication, verification, system security, etc. Thus, encryption techniques ensure the security, integrity, and confidentiality of information and prevent attacks and counterfeiting.

Along with the explosion of information technology, techniques for ensuring information security in digital communications are divided into three different types: cryptography, steganography, and watermarking. Each type of cryptographic technique has different applications and objectives, but all aim to ensure the security of confidential information transmitted over insecure channels, such as the Internet.

Cryptography and Steganography techniques are generally used to transmit sensitive information between two or multiple entities within the same group or set. However, there are differences between them.

Cryptography uses mathematical transformations to encrypt a message, turning each readable message into a random sequence of characters, called ciphertext, to be transmitted over a public network to an intended recipient. This is when two people, for example A and B, communicate with each other, even though person C cannot read the content of the information, it is clear that A and B are sending or exchanging information with each other without person C being able to grasp the content. On the other hand, with Steganography, person C cannot know if there is any secret communication or exchange of information between persons A and B [1]. To handle and ensure this important matter, persons A and B will use an intermediary object, which is digital multimedia, such as audio, video, or images…

Watermarking is similar in principle to Steganography but differs in its application purpose. The objective of Watermarking is to embed information in an image in such a way that the Watermarking cannot be shifted and does not destroy the integrity of the original image carrying the message, i.e., preserving the original content of the message during transmission. Watermarking is often applied in areas such as copyright protection.

The next part of the article will focus on cryptographic systems and their development, as well as the mathematical foundations.

2. Theoretical basis

2.1. Cryptosystem

Mathematically, a cryptosystem or encryption scheme can be defined as a tuple (P, C, K, E, D) with the following properties.

1. P is a set called the "plaintext space". Its elements are called plaintexts.
2. C is a set called the "ciphertext space". Its elements are called ciphertexts.
3. K is a set called the "key space". Its elements are called keys.
4. $E = e_k, k \in K$ is a set of functions $e_k : P \to C$. Its elements are called "encryption functions".
5. $D = d_k, k \in K$ is a set of functions $d_k : C \to P$. Its elements are called "decryption functions".

For each $e \in K$, there is $x \in K$ such that $d_x\left(e_k\left(x\right)\right) = x, \forall x \in P.$ for all .[12]

Cryptographic features. Provides a high level of security, integrity, non-repudiation, and authentication.

Confidentiality: Secure message content and data by means of encryption techniques.

Integrity: Assures the parties that the message has not been altered in transit.

Non-repudiation: Can confirm that the document has come from someone, even if they try to deny it.

Authenticity: Provides two services: identifying the origin of a message, ensuring that it is genuine. Check the identity of the person who is logging into the system, further checking their identity in case someone tries to connect and pretends to be a legitimate user**.**

2.2. RSA encryption

Symmetric encryption, for example AES, uses the same key for both encryption and decryption. The advantage of this method is fast processing speed, small size of encrypted data. However, the security is not really safe when it is necessary to exchange information at the processing level with many parties receiving and sending data. Asymmetric encryption, also known as Public Key Cryptography, is a method performed on two keys. One key is used for encryption (public key) and one key is used for decryption (private key). The private key is used to decrypt and is kept secret, the public key is the key used to generate the encryption and is made public so that anyone can use it to send messages to the subject. The decryption key cannot be calculated from the encryption key. The advantage of public encryption is that key management is more flexible and efficient. The user only needs to protect his private key. However, the disadvantage of public encryption lies in the execution speed, which is much slower than symmetric encryption.

RSA is a public key cryptographic algorithm, first described by Ron Rivest, Adi Shamir and Len Adleman in 1977 at the Massachusetts Institute of Technology (MIT). The name of the algorithm is derived from the first 3 letters of the names of the three authors [3]. The RSA algorithm was patented by MIT in the United States in 1983 (Registration No.4.405.829).

RSA is widely used in encryption and digital signature technology. In this encryption system, the public key will be shared publicly for everyone. The RSA operation is divided into 3 steps: key generation, encryption, and decryption.

Operation Description:

Public key: Open to the public and encrypted.

Private key: Publicly encrypted information can only be decrypted with the corresponding private key.

Key Generation process in RSA:

(1) Select two large prime numbers p, q with $p \neq q.$

(2) Derive the value of n as $n = p * q$.

(3) Calculate the totient value as $\phi(n) = (p-1)(q-1)$.

(4) Choose a natural number $1 < e < \phi(n)$ and is a co-prime with $\phi(n).$

(5) Calculate *d* such that *d* is congruent with *e*, i.e $de = 1 \left(\text{mod } \phi(n)\right).$

(6) Pair $\{n,e\}$ is the public key, pair $\{n,d\}$ is the private key.

Encoding/Encryption process: The sender converts the plaintext M into a number m, m < n, according to a pre-agreed reversible function (from m can redefine M). Calculate *c* as the ciphertext of m by the formula:

$c = m^e \pmod n$. Send *c* to the recipient.

Decoding/Decryption process: The receiver receives *c* and knows *n* (publicly) so it can find *m* from *c* according to the following formula: $m = c^d \pmod n$, where d is the inverse of e modulo $\phi(n)$. This inverse exists under the condition $\left(e,\phi(n)\right)=1$.

The security of the RSA system is based on two mathematical problems: the problem of prime factorization of large integers and the problem of RSA. If the above two problems are difficult, then it is not possible to perform complete decryption for RSA. The RSA problem is the problem of calculating the square root of *e* modulo *n* (where *n* is composite): find the number *m* such that $me = c \bmod n,$ where (e,n) is the public key and c is the ciphertext. Currently, the most promising method to solve this problem is to factor *n* into prime factors. When doing this, the attacker will find the secret exponent d from the public key and can decrypt it according to the algorithm's process. If an attacker finds two primes p and q such that: $n = p.q$, it is easy to find the value $(p-1)(q-1)$ and thereby determine d from e.

As of 2005, the largest number that can be factored to a prime is 663 bits with the distributed method while the RSA key has a length of 1024 to 2048 bits. Some experts believe that a 1024-bit key may soon be broken (there are also many who oppose this). With a 4096-bit key it is unlikely to be broken in the near future [1]. Therefore, it is generally assumed that RSA is secure provided that n is chosen large enough. If it is 256 bits or shorter, it can be analyzed in a few hours with a personal computer using available software. If n were 512 bits long, it could have been parsed by several hundred computers by 1999. A theoretical device called the TWIRL described by Shamir and Tromer in 2003 raised questions about security. the whole of a 1024-bit key. Therefore, it is currently recommended to use keys with a minimum length of 2048 bits.

To overcome this problem, the Extended Euclidean Algorithm (EEA) plays a very important role to generate the secret key d. Therefore, privacy is preserved. The next part of the paper presents the basic mathematical knowledge used in RSA cryptography.

2.3. Euclidean Algorithm, Extended Euclidean Algorithm

The Euclidean algorithm is a simple algorithm for calculating the greatest common divisor (GCD) of two positive integers. This algorithm is named after the Greek mathematician Euclidean in his book "Fundamentals of Geometry" (Elements).

The Euclidean algorithm works by repeatedly dividing the larger number by the smaller number, until the smaller number is zero. The application of the Euclidean algorithm is very wide in mathematics and computer science. In the RSA cryptosystem, the Euclidean algorithm is used to find a number that is relatively prime with a certain positive integer. Specifically, the process of generating public and private keys of the RSA cryptosystem is as follows:

Step 1: Choose two primes p and q with large differences, calculate n = p*q.

Step 2: Calculate co-prime with (p-1)*(q-1), is called the number e. We use Euclidean algorithm to find the number e, such that GCD(e, (p-1)*(q-1)) = 1.

Step 3: Calculate the number d such that (d*e) % ((p-1)*(q-1)) = 1. We can use the Euclidean expansion algorithm to find this number d.

Step 4: The public key is the pair (n, e), the private key is the number d.

When there is a message that needs to be encrypted, the sender uses the public key (n, e) to encrypt the message. The receiver will use the secret key d to decrypt the message. When the receiver wants to reply back to the message, they will use the public key (n, e) to encrypt the message and send it back to the sender [9].

We denote $\mathbb{Z}$ as the set of integers, $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ and $\mathbb{Z}^+$ is the set of non-negative integers, $\mathbb{Z}^+ = \{0, 1, 2, ...\}$.

The set $\mathbb{Z}$ is closed to addition, subtraction, and multiplication, but not to division: dividing an integer by an integer does not always result in an integer. Therefore, the case of divisibility, i.e. when the integer a is divided by the integer b, the quotient is an integer q, $a = b.q$, has a special meaning. Then we say that a is divisible by b, b is divisible by a, a is a multiple of b, b is a divisor of a, and denoted by $b|a$. It is easy to see right away that 1 is a divisor of all irregular integers, zero is a multiple of any integer, every integer a is a divisor, and at the same time a multiple, of itself.

Let $a, b\,(b>1)$ any two integers. By dividing a by b, we get two numbers q and r such that

$$a = b.q + r,\ 0 \le r < b.$$

The number q is called the quotient of the division of a by b, denoted $a \text{ div } b$ and the number r is called the remainder of the division a by b, the symbol $a \bmod b$.

Example: $30 \text{ div } 7 = 4$ and $30 \bmod 7 = 2, -30 \text{ div } 7 = -5$ and $-30 \bmod 7 = 5$.

An integer d is said to be a common divisor of two integers a and b if $d|a$ and $d|b$. An integer d is said to be the greatest common divisor of a and b if $d > 0$, d it is a common divisor of a and b, and all common divisors of a and b are less than or equal to d. Denote the greatest common divisor of a and b as $\gcd(a,b)$.

Example: $\gcd(12,18)=6$, $\gcd(-18,27)=9$.

It is easy to see that for every positive integer a we have $\gcd(a,0)=a$, we will also convention that gcd(0, 0) = 0.

Theorem1: *If $b \neq 0$ and $b|a$, then we have $gcd(a,b) = b$. If $a = b.q + r$, then gcd(a, b) = gcd(b, r).*

An integer m is said to be a common multiple of a and b if $a|m$ và $b|m$. The number m is called the least common multiple of a and b, and is denoted by $\text{lcm}(a,b)$, if m is a common multiple of a and b, and every common multiple of a and b is greater than or equal to m.

Example: lcm(15, 20) = 60.

any two positive integers a and b, we have the relation $\text{lcm}(a,b).\gcd(a,b)=a.b$.

From Theorem 1, we deduce the following algorithm to find the greatest common divisor of any two integers:

Euclidean algorithm to find greatest common divisor:

INPUT: two non-negative integers a and b, with $a \geq b$.

OUTPUT: greatest common divisor of a and b.

1. While b > 0, execute:

Put $r \leftarrow a \bmod b$, $a \leftarrow b$, $b \leftarrow r$.

2. Result (a) [8].

Application example 1: Using the Euclidean algorithm to find $\gcd(973, 301)$, we get the values assigned to the variables a, b and r as follows:

973 = 301.3 + 70

301= 4.70 + 21

70=3.21+ 7

21=3.7 + 0

a	b	r
973	301	21
301	70	7
70	21	0
21	7	0
7	0	

Application example 2: Using the Euclidean algorithm to find $\gcd(91470, 51066)$, we get the values assigned to the variables a, b and r as follows:

91470 = 1.51066 + 40404

51066 = 1.40404 + 10662

40404 =3.10662+ 8418

10662 =1.8418 + 2244

8418 = 3.2244 + 1686

2244 = 1.1686 + 558

1686 = 3.558 + 12

558 = 46.12 + 6

12 = 2.6 + 0

a	*b*	*r*
91470	51066	40404
51066	40404	10662
40404	10662	8418
10662	8418	2244
8418	2244	1686
2244	1686	558
1686	558	12
558	12	6
12	6	0
6	0	0

Extended Euclidean Algorithm:

This algorithm aims to determine 3 integers *x, y,d* such that: where *m,n* are two given integers with the assumption $m \geq n$. The content of the algorithm is as follows:

Given 3 vectors $(a_1,a_2,a_3),(b_1,b_2,b_3),(c_1,c_2,c_3)$; steps are as follows:

Step 1. $(a_1,a_2,a_3) \leftarrow (1,0,m),(b_1,b_2,b_3) \leftarrow (0,1,n)$;

Step 2. If $b_3 = 0$ then the algorithm stops and (a_1,a_2,a_3) is the answer;

Step 3. Put $q = \left[\frac{a_3}{b_3}\right]$; và $(c_1,c_2,c_3) \leftarrow (a_1,a_2,a_3) - q.(b_1,b_2,b_3); (a_1,a_2,a_3) \leftarrow (b_1,b_2,b_3); (b_1,b_2,b_3)$ $\leftarrow (c_1,c_2,c_3)$ and go to step 2.

Application example 1: Using the extended Euclidean algorithm for numbers a = 42823 and b = 6409, we get the following values for variables $a,b,q,r,x,y,x_1,x_2,y_1,y_2$ respectively (after each cycle of executing two instructions 2 and 3):

a	*b*	*q*	*r*	*x*	*y*	x_1	x_2	y_1	y_2
42823	6409					0	1	1	0
6409	4369	6	4369	1	-6	1	0	-6	1
4369	2040	1	2040	-1	7	-1	1	7	-6
2040	289	2	289	3	-20	5	-1	-20	7
289	17	7	17	-22	147	-22	3	147	-20
17	0	17	*0*	377	-2519	377	-22	-2519	147

Executing the cycle of two instructions 2 and 3, the obtained x,y,r values always satisfy $42823.x + 6409.y = r$, and so at the end of the loops (corresponding to the value $b = 0$), we get the

result $d=17, x=-22$ and $y=147$, pairs of numbers $(-22,147)$ satisfy: $42823.-22+6409.147=17.$

Application example 2: Using the extended Euclidean algorithm for numbers $a=1970$ and $b=1066,$ we get the following values for variables $a,b,q,r,x,y,x_1,x_2,y_1,y_2$ respectively (after each cycle of executing two instructions 2 and 3):

a	*b*	*q*	*r*	*x*	*y*	x_1	x_2	y_1	y_2
1970	1066					0	1	1	0
1066	904	1	904	1	-1	1	0	-1	1
904	162	1	162	-1	2	-1	1	2	-1
162	94	5	94	6	-11	6	-1	-11	2
94	68	1	68	-7	13	-7	6	13	-11
68	26	1	26	13	-24	13	-7	-24	13
26	16	2	16	-33	61	-33	13	61	-24
16	10	1	10	46	-85	46	-33	-85	61
10	6	1	6	-79	146	-79	46	146	-85
6	4	1	4	125	-231	125	-79	-231	146
4	2	1	2	204	377	-204	125	377	-231
2	0	2	0	533	-985	533	-204	-985	377

Executing the cycle of two instructions 2 and 3, the obtained x,y,r values always satisfy $1970.x+1066.y=r,$ and so at the end of the loops (corresponding to the value $b=0$), we get the result $d=2, x=-204$ and $y=377,$ pairs of numbers $(-204,377)$ satisfy: $1970.-224+1066.377=2.$

3. Applying Euclidean extension algorithm in RSA encryption

Key Generation process in RSA:

(1) Select two large prime numbers p,q with $p\neq q.$

(2) Derive the value of n as $n=p*q$.

(3) Calculate the totient value as $\phi(n)=(p-1)(q-1)$.

(4) Choose a natural number $1<e<\phi(n)$ and is a co-prime with $\phi(n)$.

(5) Compute '*d*' by using Extended Euclidean Algorithm [6].

Extended Euclidean Algorithmto find d:

The Extended Euclidean Algorithm has the equation as $\phi(n).x + e.y = \gcd(\phi(n), e)$ where $\phi(n), e$ are generated as shown above. In order to find the values of *x, y* the steps are as follows:

1. **Table 1.**

Row	a	b	$\phi(n)$	e
1	1	0	$\phi(n)$	-
2	0	1	e	e_1
3	a_3	b_3	d_1	e_2
⋮	⋮	⋮	⋮	⋮
i	a_i	b_i	d_{i-2}	e_{i-1}
⋮	⋮	⋮	⋮	⋮
k	$a_k \to x$	$b_k \to y$	1	e_{k-1}

2. A table is constructed comprising the values of a, b, $\phi(n)$, e is shown below:

The values of $a_3, a_4, ..., a_k$ is computed by using equation $a_n = a_{\lceil (n-1)-1 \rceil} - a_{(n-1)}.e_{\lceil (n-1)-1 \rceil}$.

The values of $b_3, b_4, ..., b_k$ is computed by using equation $b_n = b_{\lceil (n-1)-1 \rceil} - b_{(n-1)}.e_{\lceil (n-1)-1 \rceil}$.

3. Derive d_n as $d_n = d_{\lceil (n-1)-1 \rceil} - d_{(n-1)}.e_n$, conventions $d_{-1} = \phi(n), d_0 = e.$

4. The value of e_n is obtained by $e_n = \left\lceil \frac{d_{\lceil (n-1)-1 \rceil}}{d_{(n-1)}} \right\rceil$.

5. The table is formulated until $\phi(n)$ becomes 1.

6. From the above table the values of x and y are obtained as shown.

7. Substitute the values of x, y in the given equation: $\phi(n).x + e.y = \gcd(\phi(n), e)$ which yields 'd' value, known as private key.

8. The conditions for selecting ‘d’ value are:

+ If $d > \phi(n)$, then $d = d \bmod \phi(n)$.

+ If $d < 0$, then $d = d + \phi(n)$. The generated ‘d’ value is the private key for decryption process.

Application example 1: Given $e = 17; \phi(n) = 3120$, compute the private key 'd'.

Row	a	b	$\phi(n)$	e
1	1	0	3120	-
2	0	1	17	183

3	1	-183	9	1
4	-1	184	8	1
5	2	-376	1	8

Ta có: $2.3120 + (-367).17 = \gcd(3120,17) = 1$. Vì d < 0 nên $d = -367 + 3120 = 2753.$

Application example 2: Given $e = 3491; \phi(n) = 978072$, compute the private key 'd'.

Row	a	b	$\phi(n)$	e
1	1	0	978072	-
2	0	1	3491	280
3	1	=280	592	5
4	-5	1401	531	1
5	6	-1681	61	8
6	-53	14849	43	1
7	59	-16530	18	2
8	-171	47909	7	2
9	401	-112348	4	1
10	-572	160257	3	1
11	973	-272605	1	3

We have: $973.978072 + (-272605).3491 = \gcd(978072, 3491) = 1.$

Since d < 0, we obtain $d = -272605 + 978072 = 705467.$

4. Conclusion

In cryptography, the Euclidean algorithm is used to compute the public and private keys in RSA-based cryptosystems. This paper presents the basic knowledge of RSA cryptosystem, Euclidean algorithm and the use of extended Euclidean algorithm to find secret keys with large data.

References

[1]. D. R. Stinson , Cryptography_ Theory and Practice, USA: CRC Press,1995.

[2]. K. Singh, R. Verma, R. Chehal , Modified Prime Number Factorization Algorithm (MPFA) For RSA Public Key Encryption, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, (2012) pp 204-206.

[3]. M. T. Rhee, Cryptography and Secure Communications, USA: McGraw - Hill Book Co,1994.

[4]. Ph.D. W. Stallings , Network and Internetwork Security Principles and Practice, USA: Prentice Hall,1995.

[5]. S. Sharma, J.- S. Yadav, P. Sharma, Modified RSA Public Key Cryptosystem Using Short Range Natural Number Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering, (2012) Volume 2,Issue 8, pp 134-138.

[6] J. Zhou, J. Hu, P. Chen The 2nd , Extended Euclid algorithm and its application in RSA, International Conference on Information Science and Engineering, IEEE, (2010) 10.1109/ICISE.2010.5691644, 4-6

[7] R Gennaro, RSA-Based Undeniable Signatures, Journal of Cryptology, (2000), Vol 13, No. 4, pp 397-416.

[8]. P.H. Đien, Information coding - Mathematical foundations and applications, Institute of Mathematics, VAST, Ha noi, 2004.

[9]. P.H. Đien, H. H. Khoai, Algorithmic Arithmetic - Theoretical Foundations & Practical Calculations, Institute of Mathematics, VAST, Ha noi, 2002.

[10]. Gennaro, Robust and Efficient Sharing of RSA Functions, Journal of Cryptology, (2008) Vol 13, No 2, pp 273-300.

[11]. G. Shahi, C.Singh, Cryptography and its two Implementation Approaches, International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 1, Issue 3, May 2013,PP 668-672.

[12]. Buchmann, A. Johannes, Introduction to Cryptography (2nd ed.). Springer. ISBN 0-387-20756-2,13 July 2004.

[13]. R. Rivest, A. Shamir, L. Adleman, A Method for Obtaining Digital Signatures and Public-Key Cryptosystems, Communications of the ACM 21 (2): 120–126, doi: 10.1145/359340.359342, 1977