Factorization Hack of RSA Secret Numbers

Andysah Putera Utama Siahaan
Faculty of Computer Science
Universitas Pembangunan Panca Budi
Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing, 20122, Medan, Sumatera Utara, Indonesia

***Abstract* -** *RSA always uses two big prime numbers to deal with the encryption process. The public key is obtained from the multiplication of both figures. However, we can break it by doing factorization to split the public key into two individual numbers. Cryptanalysis can perform the public key crack by knowing its value. The private key will be soon constructed after the two numbers retrieved. The public key is noted as "N", while "N = P . Q". This technique is unclassified anymore to solve the RSA public and private key. If it is successfully factored into p and q then ϕ (N) = (P -1). (Q -1) can be further calculated. By having the public key e, the private key d will be solved. Factorization method is the best way to do the demolition. This study concerns to numbers factorization. GCD calculation will produce the encryption "E" and decryption "D" keys, but it depends on the computer speed.*

***Keywords* -** *Cryptography, RSA, Public Key, Factorization*

I. Introduction

There are various techniques should be used to protect the confidential image data from unauthorized accessand there are many to breach the security as well[1]. RSA is a public key cryptographic algorithm that works on two main cryptographic processes, public and private key [6].The power of RSA is in the prime numbers. The longer the key is used, the longer the time used to factor the public key. The main strength of the RSA algorithm is based on the difficulty of factoring large numbers into prime factors [3]. Although RSA is still difficult to solve, it has a security hole by simply knowing the ciphertext and public key. One of the techniques is to attack the RSA public key by factoring the "N". This weakness is utilized to carry out attacks to test the security level of this algorithm.The most common attack on RSA is the factorization problem of handling enormous number. If there is a new rapid method has been developed, it is possible to dismantle the RSA. In 2005, the largest number factorization was commonly used throughout the 663 bits, using advanced distribution methods. RSA keys in general throughout 1024 through 2048 bits. Some experts believe that 1024-bit keys will be solved shortly, but 2048-bit keys are still difficult to solve in the future.Since RSA is breakable by factoring the "N", the security of RSA is often based on the integer factorization problem [2].According to mathematical theory, it is easy to get two big prime numbers. However, the factorization is not easy as we think [5]. We try to find the value of "P" and "Q" form the known public key, but we limit the length of the "N" value according to the computer clock speed.

II. Basic Concept

The symmetric key is one of the cryptographic systems that uses the same kind of keys for encryption and decryption while the asymmetric is where the encryption and decryption use the different keys [4]. The RSA algorithm is described by three people from MIT (Massachusetts Institute of Technology). There are Ron Rivest, Adi Shamir, and Len Adleman in 1977. The encryption and description processes on RSA come from the concept of prime numbers, and modulo arithmetic. The encryption and decryption keys are both integers. The encryption key is unclassified while the decrypt key is confidential. The decryption key is generated from several pieces of prime numbers together with the encryption key. This algorithm patented by MIT in 1983 in the United States as US Patent 4405829. This patent is valid until September 21st, 2000.

RSA has several attributes indicate the input and output parameters. There are seven aspects we have to understand before trying to hack the RSA secret key such as:

1.	P and Q	:	private
2.	N = P . Q	:	public
3.	φ(N) = (P -1)(Q -1)	:	private
4.	E (encryption key)	:	public
5.	D (decryption key)	:	private
6.	M (plaintext)	:	private
7.	C (ciphertext)	:	public

There are two signs above, public and private. Public means the value is unclassified. Everybody can get the information. However, the private means it is confidential. We must save it secretly. We hope nobody will find those value. The "N" value is obtained from the multiplication "P" and "Q". Euler function is a function that is used in mathematical calculations on the RSA algorithm. Euler function is a function that is used in mathematical calculations on the RSA algorithm. Euler function defines ϕ (N) for N≥ 1 indicates the number of positive integers <N

relatively prime to N. Two numbers "A" and "B" are relatively prime if GCD(A, B) = 1. The value of "E" and "D" are used for both encryption and decryption respectively. The "M" and "C" are the plaintext and ciphertext byte arrays.

III. Evaluation

Implementation of RSA describes how the algorithm performs the message concealment process until the message is hidden. There are three most important stages of manipulating the message in the application of the RSA algorithm, such as the key generation process, the process of concealment of random messages (encryption) and the process of restoring the random messages into the initial message before manipulation (decryption). To hack the RSA secret key, we can simulate from the first step, that is the key generation process. After the generator produces it, the computer will note the values written on screen.

Table 1. Example of Key Generation

Variable	Value
P	5011
Q	1093
N	5477023
φ	5470920

Table 1 illustrates the form of the key generation. The computer produces the random numbers for both "P" and "Q". Then "N" and "φ" are calculated by the earlier formula. The encryption and decryption keys can be generated by both "N" and "φ". Then we carry out the GCD between the"φ"and new incremented variable "E". Not all the variable works. The test of GCD(E, φ) = 1 must be done. The value which GCD ≠1 cannot be accepted. We check the "E" from 2 to 30 and the result is that we have five possibilities, such as 11, 17, 19, 23, and 29. Sometimes, it is hard to guess which value is used by the sender since there are many probabilities of the prime numbers from 2 to φ. In this example, we use φ=540920. There are many prime numbers from 2 until 540920. However, it will be discovered. It depends on the computer speed. In this work, we assume that the value of the encryption used is 11. Tabel 2 shows the selected values which GCD is 1.

Table 2. The Encryption Key and GCD

E	GCD
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	1
12	12
13	13
14	14
15	15
16	8
17	1
18	18
19	1
20	20
21	21
22	2
23	1
24	24
25	5
26	26
27	9
28	28
29	1
30	30

Soon after the encryption key is obtained, the decryption key can be processed by doing the following formula.

$$D = \frac{\varphi.\mathrm{K} + 1}{E} \qquad (1)$$

$$R = (\varphi.\mathrm{K} + 1)\ Mod\ E \qquad (2)$$

The first formula is to find the "D" (encryption value). But not all the value retrieved can be used. The second is Rest, "R" is a determinant whether the value of "D" works. Looping is very important to search for the "R" return zero value. Since R has more than zero, "D" will not work as the decryption key.

Table 3. The decryption key

K	D	Rest
2	994712,82	9
3	1492069,2	2
4	1989425,5	6
5	2486781,9	10
6	2984138,3	3
7	3481494,6	7
8	3978851	0
9	4476207,4	4
10	4973563,7	8
11	5470920,1	1
12	5968276,5	5

13	6465632,8	9
14	6962989,2	2
15	7460345,5	6
16	7957701,9	10
17	8455058,3	3
18	8952414,6	7
19	9449771	0
20	9947127,4	4
21	10444484	8
22	10941840	1
23	11439196	5
24	11936553	9
25	12433909	2
26	12931266	6
27	13428622	10
28	13925978	3
29	14423335	7
30	14920691	0

Table 3 shows there are 29 sample of “K”, “D” and “Rest”. There are only three datas that states zero value, such as data 8, 19 and 30. But, in this case, the correct decryption key for the encryption key 11 as discussed before is the first discovery, that is K = 8 and D = 3978851.In Table 4, all parameters can directly implemented to plaintext and ciphertext.

Table 4. The complete parameters of RSA

Variable	Value
P	5011
Q	1093
N	5477023
T	5470920
E	11
D	3978851

That was the complete story of RSA. In this section, we try to break the N = 5477023. By factorizing it, we will obtain the “P” and “Q” simultenaously. First we calculate the SQRT(N) and test the value is square root or not.

A = SQRT(N)
= SQRT(5477023)
= 2340.304 (must be rounded down)
= 2340

Since A . A ≠ N, 5475600 ≠ 5477023, we add A to 1. The value of “A” will be 2341

A = A + 1
= 2340 + 1
= 2341

B = A . A – N
= 2341. 2341- 5477023
= 3258

C = $\sqrt{B}$
= $\sqrt{3258}$
= 57

D = C . C
= 57 . 57
= 3249

R = B – D
= 3258 - 3249
= 9

The variable “R” is used to determined the two prime numbers. “R” means Rest. The progress is looped until the R=0 is achieved. Since “R” does not produce zero, the “A” and “B” are malfunctioned. In Table 5 below, the complete progress of these five parameters cycle.

Table 5. The cycle of “A”, “B”, ”C”, “D” and “R” (part 1)

Cycle	A	B	C	D	R
1	2341	3258	57	3249	9
2	2342	7941	89	7921	20
3	2343	12626	112	12544	82
4	2344	17313	131	17161	152
5	2345	22002	148	21904	98
6	2346	26693	163	26569	124
7	2347	31386	177	31329	57
8	2348	36081	189	35721	360
9	2349	40778	201	40401	377
10	2350	45477	213	45369	108
11	2351	50178	224	50176	2
12	2352	54881	234	54756	125
13	2353	59586	244	59536	50
14	2354	64293	253	64009	284
15	2355	69002	262	68644	358
16	2356	73713	271	73441	272
17	2357	78426	280	78400	26
18	2358	83141	288	82944	197
19	2359	87858	296	87616	242
20	2360	92577	304	92416	161
21	2361	97298	311	96721	577
22	2362	102021	319	101761	260
23	2363	106746	326	106276	470
24	2364	111473	333	110889	584
25	2365	116202	340	115600	602
26	2366	120933	347	120409	524
27	2367	125666	354	125316	350
28	2368	130401	361	130321	80
29	2369	135138	367	134689	449
30	2370	139877	374	139876	1

Let's focus on the "R" value. We have done the calculation for 30 times. However, the "R" still return the wrong value. The progress must be looped until it shows zero.

Table 6. The cycle of "A", "B", "C", "D" and "R" (part 2)

Cycle	A	B	C	D	R
683	3023	3661506	1913	3659569	1937
684	3024	3667553	1915	3667225	328
685	3025	3673602	1916	3671056	2546
686	3026	3679653	1918	3678724	929
687	3027	3685706	1919	3682561	3145
688	3028	3691761	1921	3690241	1520
689	3029	3697818	1922	3694084	3734
690	3030	3703877	1924	3701776	2101
691	3031	3709938	1926	3709476	462
692	3032	3716001	1927	3713329	2672
693	3033	3722066	1929	3721041	1025
694	3034	3728133	1930	3724900	3233
695	3035	3734202	1932	3732624	1578
696	3036	3740273	1933	3736489	3784
697	3037	3746346	1935	3744225	2121
698	3038	3752421	1937	3751969	452
699	3039	3758498	1938	3755844	2654
700	3040	3764577	1940	3763600	977
701	3041	3770658	1941	3767481	3177
702	3042	3776741	1943	3775249	1492
703	3043	3782826	1944	3779136	3690
704	3044	3788913	1946	3786916	1997
705	3045	3795002	1948	3794704	298
706	3046	3801093	1949	3798601	2492
707	3047	3807186	1951	3806401	785
708	3048	3813281	1952	3810304	2977
709	3049	3819378	1954	3818116	1262
710	3050	3825477	1955	3822025	3452
711	3051	3831578	1957	3829849	1729
712	3052	3837681	1959	3837681	0

The calculation continues until reach the desired value. Table 6 is the last cycle before reaching the zero value. After cycle 712 the "R" value gave the right value. R = 0 is what the factorization needs. The "P" and "Q" can be obtained then.

$$\begin{aligned} P &= A - \sqrt{B} \\ &= 3052 - \sqrt{3837681} \\ &= 3052 - 1959 \\ &= 1093 \end{aligned}$$

$$\begin{aligned} Q &= A + \sqrt{B} \\ &= 3052 + \sqrt{3837681} \\ &= 3052 + 1959 \\ &= 5011 \end{aligned}$$

Now we see the P = 1093 and Q = 5011, and the values before factorization are P = 5011 and Q = 1039. It does not matter because the "P" and "Q" are swappable.After we get those values, the process of finding "E" and "D" can be continued as discussed before.

IV. Conclusion

There are many algorithms for symmetric and asymmetric can be used for encryption, decryption, key exchange and digital signature. To resolve the prime factors of RSA, we can use integer factorization algorithm. The RSA breach can be resolved easily by doing factorization on the public key. Everyone has already known the issue, and it is not a secret thing anymore. The speed depends on the public key length. The value of "P" and "Q" are not confidential since we use the small integer number. To protect them, we should use the big integer. There is no limitation for the integer. We can increase as long as the computer can calculate the formula properly.

References

[1] A. P. U. Siahaan, "RC4 Technique in Visual Cryptography," *SNATIA*, Semarang, 2016.

[2] B. R. Ambedkar dan S. S. Bedi, "A New Factorization Method to Factorize RSA Public Key Encryption," *International Journal of Computer Science*, vol. 8, no. 6, pp. 242-247, 2011.

[3] H.-M. Sun, M.-E. Wu, W.-C. Ting dan M. J. Hinek, "Dual RSA and Its Security Analysis," *IEEE Transactions on Information Theory*, vol. 53, no. 8, 2007.

[4] A. P. U. Siahaan, "Three-Pass Protocol Concept in Hill Cipher Encryption Technique," *SNATI*, Yogyakarta, 2016.

[5] N. Hegde dan P. Deepthi, "Pollard RHO Algorithm for Integer Factorization and Discrete Logarithm Problem," *International Journal of Computer Applications*, vol. 121, no. 18, pp. 14-17, 2012.

[6] F. Nizar, F. Latheef dan A. Jamal, "RSA Based Encrypted Data Embedding Using APPM," *International Journal of Engineering Trends and Technology*, vol. 9, no. 15, pp. 777-782, 2014.