A Probabilistic Encryption Way in Knapsack PKC

Zuxi Wang
Science and Technology on Multi-spectral Information Processing Laboratory
Huazhong University of Science & Technology
Wuhan, P. R. of CHINA, 430074
zuxiw@mail.hust.edu.cn

Baimu Yu
Science and Technology on Multi-spectral Information Processing Laboratory
Huazhong University of Science & Technology
Wuhan, P. R. of CHINA, 430074
yubaimu@163.com

***Abstract*—with the proposed of quantum computers, the knapsack public-key cryptosystem (PKC) becomes very popular. At present the most significant challenge towards knapsack PKC is lattice attack, and density is one of the most important factors to measure the security of it. In this paper, we introduce a probabilistic way to encrypt and compared the knapsack quadratic knapsack and the probabilistic knapsack.. We know that the probabilistic scheme enjoys a high density, information rate and it also secure against brute force attack, statistical analysis attack, lattice attack and so on, and it is efficient and practical.**

Keywords-Probabilistic encryption; Statistical analysis attack; Lattice basis reduction

I. INTRODUCTION

Since Diffie-Hellman proposed public-key cryptosystem (PKC) in "New directions in cryptography"[2], public-key cryptosystem becomes very popular. Lots of classical algorithms like RSA and ECC appeared. With the proposing of quantum computer and the enhancement of the computer calculation ability, the computer can solve many mathematic problems[13]. So the designers are actively to find a new cryptosystem to cope with the challenge of the quantum computer. In 1998, Stony Brook University algorithm database showed that knapsack is the one of the hottest research in the 75 kinds of algorithms[10], and quantum computer can't solve this kind of combinatorial optimization problems[14].

The first knapsack PKC which is proposed by Markle-Hellman is a super increasing knapsack encryption scheme[3]. Lagarias and Odlyzko proposed the low-density attack (LDA) for solving general low-density subset sum problems[6]. Coster et al had strictly proved that a knapsack which density is less than 0.9408 will suffer from this attack[8]. So density plays an important role in the security of the knapsack PKC.

In order to make the density high, the CR scheme[4] and the OTU scheme[5] use a low weight encoding . However, Nguyen and Stern showed that the low weight encoding is not secure[7]. So many designers prefer to use quadratic knapsack to achieve a very high density. The first quadratic knapsack was proposed by Gallo et al[9]. They can easily solve the density problem. And the quadratic knapsack is a nonlinear PKC. So it is much safer than the linear system. But it also has some problems in encryption. It has to keep the quadratic residue table as secret key which will increase the difficulty of key management.

So in the section 2, we give some background knowledge. In section 3, we give a new knapsack encryption way. In section 4 we compare the knapsack, quadratic knapsack and the new knapsack and know that the new knapsack scheme is efficient and practical

II. BACKGROUND KNOWLEDGE

A. *Knapsack problem*

The knapsack problem is to find the solution $(x_1,\cdots,x_n)\in I,\ I=\{0,\cdots,p-1\}$ that satisfied the linear Diophantine equation

$$D=a_1x_1+a_2x_2+\cdots+a_nx_n \qquad (1)$$

for giving positive integers $a_1,a_2,\cdots,a_n$ and *C*. To solve this Diophantine equation is to seek integer solution of $(x_1,x_2,\cdots,x_n)$ over an integer ring. And the result is not unique.

Another problem is quadratic knapsack problem or matrix cover problem which has to solve the Diophantine equation

$$D=a_1x_1^2+a_2x_2^2+\cdots+a_nx_n^2 \qquad (2)$$

where $x_i\in I$ and $a_i, i=1,\cdots,n$ are positive integers.

These problems are used to construct knapsack PKC[16-18]. In this paper we compared these two knapsacks, and find some relationship about it.

B. *Definition of Lattice*

A lattice is a discrete subgroup of R^n . In another words, every lattice *L* is generated by some set of linearly independent vectors $\{v_i\}\in L$, called a basis of *L*, i.e.

$$L=\{z_1v_1+\cdots+z_nv_n \mid z_1,\cdots,z_n\in Z\}$$

The most famous problem in lattices theory includes the shortest vector problem(SVP), the closest vector problem (CVP) and the smallest basis problem(SBP). SVP is to find the shortest vector in L. CVP is to minimize the length of the length of the vector s-v in L, where s and L are given. The best polynomial-time algorithms for solving SVP are the LLL algorithm[12]. The shortest lattice is the solution vector of the low density subset-sum problem on a great probability. And no more polynomial time algorithm can solve these three problems. So these problems have significance in complexity theory and cryptography.

C. *Definition of Density*

There are different definitions when the knapsacks are based on different problems, such as the density of 0-1 knapsack[6], the density of non-0-1 knapsack[11] and the pseudo-density of quadratic knapsacks[7][15].

In this paper, we use the definition in[11]:

$$d = \frac{nm}{\log_2(C_{\max})}$$

Where, Cmax stands for the maximum of the ciphertext, m stands for the binary length of mi.

Coster et al strictly proved that a knapsack which density is less than 0.9408 can be broken by lattice algorithms[12] Here we take 0.9408 as the standard of the security of the system.

III. A NEW KNAPSACK ENCRYPTION WAY

A. *A new way to encrypt*

Knapsack and the quadratic knapsack public-key model is traditional model which is described in Fig 1.

Plaintext
Encryption Black box
Ciphertext

Figure 1. traditional encryption model

This model is a traditional one that one plaintext corresponds to one ciphertext. The attackers can easily get the relationship between the plaintext and the ciphertext. So it is not secure against the statistical analysis attack.

Inspired from the one time pad which has a good performance in security, we add some random number to the system. So the ciphertext are different even then plaintext is the same. We can easily break the relationship between the plaintext and the ciphertext. The new model is as follows.

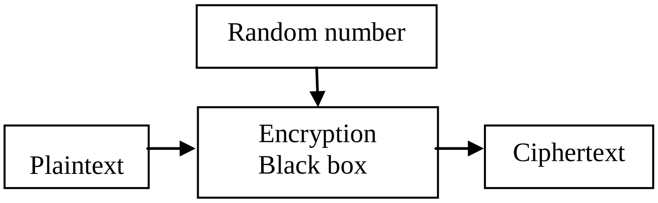

Figure 2. the new encryption model

From Fig 2, we know that the ciphertext not only have relation with the plaintext but also have relationship with random number. It is more complex than the traditional one. And the system will enjoy a very high density by adding some random number to the system. The detail security analysis shown in section 4

B. *encryption form*

The new encryption form:

$$C = \varepsilon + \sum_{i=1}^{h} a_i m_i + \sum_{i=h+1}^{n} a_i \varepsilon m_i \mod N \qquad (3)$$

$a_1, \cdots, a_n, N$ are public-keys. ε is a random number with binary length e. i.e. $|\varepsilon|_2 = e$ and does not need to be transferred. *C* is the ciphertext which need to be transferred.

IV. COMPARISON

In fact, the properties of the system have relationship with not only encrypt form but also key generation. In order to describe the advantage of the new encryption form better, we introduce a specific algorithm.

A. *Knapsack algorithm*

1) Key generation

As space is limited, here we simply describe the key generation. All the symbols are the same with [1], if the reader wants to know the details, please refers to [1].

1. Construct two simple knapsack vectors

$A = (a_1, ..., a_n), B = (b_1, ..., b_n)$

2. Randomly choose a 2-dimensional matrix C

3. Compute $\begin{pmatrix} \hat{A} \\ \hat{B} \end{pmatrix} = \begin{pmatrix} \hat{a}_1 & \cdots & \hat{a}_n \\ \hat{b}_1 & \cdots & \hat{b}_n \end{pmatrix} = C \begin{pmatrix} A \\ B \end{pmatrix}$

4. Randomly choose two prime integers *p* and *q*, and compute $N = pq$

5. Use Chinese remainder theorem to generate a cargo vector $E = (e_1, \cdots, e_n),\ e_i = \hat{a}_i (\mathrm{mod}\ p),\ e_i = \hat{b}_i (\mathrm{mod}\ q)$

6. Randomly choose an invertible integer v over Z_N.

7. Compute $F = (f_1, \cdots, f_n)$, $f_i = e_i v \mod N$.

Public key: *F*.

Secret key: N, p, q, C^{-1}, v^{-1}

2) Decryption

In this part, we just introduce how to decrypt when encrypt as (3). If the readers want to know how the quadratic knapsack works, refers to [1]. If the readers want to know how the knapsack decrypts, he can let $\varepsilon = 1$.

1. Compute $t = cv^{-1} \equiv \sum_{i=1}^{n} e_i m_i (\mathrm{modN})$

2. Compute $t_p = t \bmod p, t_q = t \bmod q$

$(\mathrm{s_A}, \mathrm{s_B})^{\mathrm{T}} = C^{-1} (\mathrm{t_p}, \mathrm{t_q})^{\mathrm{T}}$

3. Compute $c_i = \gcd(\mathrm{a_n}, ..., \mathrm{a_{n-i+1}}), d_i = \gcd(\mathrm{b_n}, ..., \mathrm{b_{n-i+1}})$

$G = \{g_i = (\mathrm{g_{1i}}, \mathrm{g_{2i}}) = (\mathrm{c_{i-1}}/\mathrm{c_i}, \mathrm{d_{i-1}}/\mathrm{d_i}) \mid \mathrm{i} = 1, ...\mathrm{n}\}$

4. Compute $x_1 = (s_A, s_B) \bmod g_n$

5. When *i=1,…, h*, compute

$$r_{1i}=s_A\text{-}\varepsilon\text{-}\sum_{j=1}^{i-1}a_jm_j,r_{2i}=s_B\text{-}\varepsilon\text{-}\sum_{j=1}^{i-1}b_jm_j$$

Then $m_i=\left(\left(\frac{a_i}{c_{n\text{-}i+1}}\right)^{-1}\frac{r_{1i}}{c_{n\text{-}i+1}},\left(\frac{b_i}{d_{n\text{-}i+1}}\right)^{-1}\frac{r_{2i}}{d_{n\text{-}i+1}}\right)\bmod g_{n\text{-}i+1}$

When *i=h+1,...,n-1*, compute

$$\begin{cases}r_{1i}=(s_A\text{-}\varepsilon\text{-}\sum_{j=1}^{h}a_jm_j)/\varepsilon-\sum_{j=h+1}^{i-1}a_jm_j\\ r_{2i}=(s_B\text{-}\varepsilon\text{-}\sum_{j=1}^{h}b_jm_j)/\varepsilon-\sum_{j=h+1}^{i-1}b_jm_j\end{cases}$$

Then $m_i=\left(\left(\frac{a_i}{c_{n\text{-}i+1}}\right)^{-1}\frac{r_{1i}}{c_{n\text{-}i+1}},\left(\frac{b_i}{d_{n\text{-}i+1}}\right)^{-1}\frac{r_{2i}}{d_{n\text{-}i+1}}\right)\bmod g_{n\text{-}i+1}$

When *i=n*, compute

$$m_n=[(s_A\text{-}\varepsilon\text{-}\sum_{j=1}^{h}a_jm_j)/\varepsilon-\sum_{j=h+1}^{n-1}a_jm_j]/a_n$$

B. Comparison and analysis

1)Density and information rate

Though using the same key generation, these three kinds of encryption schemes also have differences in condition they have to satisfy.

- The knapsack must satisfy

$$p_1>\sum_{i=1}^{n}\hat{a}_ix_i,q_1>\sum_{i=1}^{n}\hat{b}_ix_i\ ,$$

- The quadratic knapsack must satisfy

$$p_2>\sum_{i=1}^{n}\hat{a}_iy_i^2,q_2>\sum_{i=1}^{n}\hat{b}_iy_i^2\ .$$

- The new knapsack scheme must satisfy

$$\begin{cases}p_3>\varepsilon+\sum_{i=1}^{h}\hat{a}_im_i+\sum_{i=h+1}^{i=h+1}\hat{a}_i\varepsilon m_i\\ q_3>b\varepsilon+\sum_{i=1}^{h}\hat{b}_im_i+\sum_{i=h+1}^{n}\hat{b}_i\varepsilon m_i\end{cases}$$

We compute the density and the information rate of these three systems separately. The relationship shows in Fig 3 and Fig 4. Here we choose n=50.

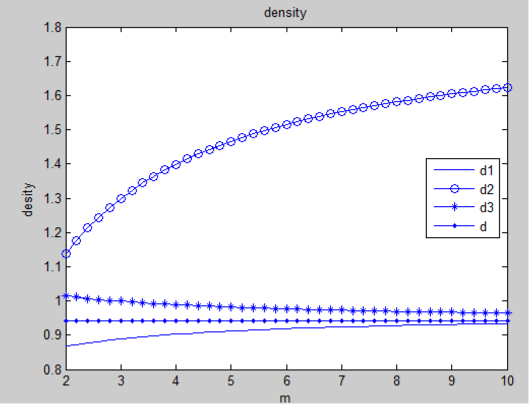

Figure 3. the relationship of density

In Fig 3, the dotted line represents the standard of density. Above the dotted line, the density is greater than 0.9408. Below the dotted line, the density is less than 0.9408. d1 represent the density of knapsack, d2 represent the density of quadratic knapsack. d3 represent the new knapsack scheme. From Fig 3 we know that the density of the new knapsack and the quadratic knapsack can be easily larger that 0.9408, but the knapsack PKC is hard to be up to 0.9408. So the quadratic knapsack and the new knapsack are secure against Shamir Secret key recovery attack and Low density subset-sum attack.

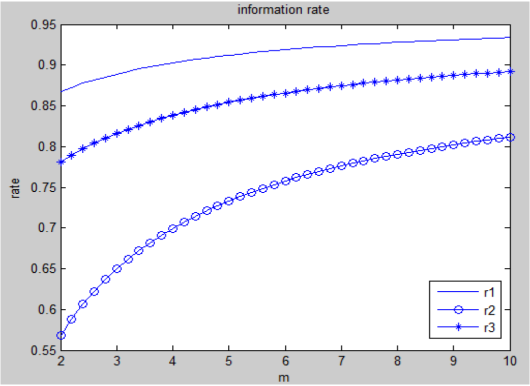

Figure 4. the relation of information rate

In Fig 4, *r1* represent the information rate of knapsack. *r2* represent the information rate of quadratic knapsack. *r3* represent the information rate of the new knapsack scheme. From the Fig 4, we know that the new knapsack scheme is very large relatively, and the quadratic is much smaller than the new knapsack scheme. So the new scheme can have very high space utilization.

2)Brute force attack

A brute way to break the system is to find $M=(m_1,m_2,\cdots m_n)$ which satisfied (3). The attackers can exhaust $\sum_{i=1}^{h}a_im_i+\sum_{i=h+1}^{n}a_i\varepsilon m_i+\varepsilon \bmod N$ for $m_i\in I$. This attack needs at least $np^{n/2}$ steps.

Suggest that the $p=m$ in the knapsack PKC, then $p=m/2$ in the quadratic knapsack PKC, and $p=m$ in the new knapsack way. So we know that the probability of the broken the quadratic knapsack is $2^{n/2}$ larger than the new way. The new knapsack is much safer than the quadratic systems.

3)Probability properties

Only the new knapsack has the probability properties. The parameter ε has greatly impact on the result of encryption. From (3) we know that the ciphertext is depended on not only the plaintext and the parameter ε . It is for this reason that the ciphertext are differed even the plaintext is same. So the distribution of the ciphertext is not the same with the plaintext. So it is secure against the Statistical analysis attack.

4)Lattice attack

Lattice basis reduction is one of the most important content in the lattice theory and it is also one of the most important tool in design and analysis. In the theory research, many lattice problems can be solved with lattice basis reduction. In cryptography, the attackers can get the plaintext when he just knows the ciphertext and the encryption function by using lattice basis reduction.

From (3), we know that there is a k that satisfied the equation $\sum_{i=1}^{h} f_i m_i + \sum_{i=h+1}^{n} f_i m_i \varepsilon + \varepsilon + kN = c$. The attackers can construct matrix V

$$V = \begin{bmatrix} 1 & \cdots & 0 & 0 & 0 & f_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & f_n \\ 0 & \cdots & 0 & 1 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 1 & N \\ 0 & \cdots & 0 & 0 & 0 & -c \end{bmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \\ v_{n+1} \\ v_{n+2} \\ v_{n+3} \end{pmatrix}$$

The linear combination of these vectors is a lattice $L = \left\{ z_1 v_1 + \cdots + z_{n+3} v_{n+3} \middle| z_1, \cdots z_{n+3} \in Z \right\}$ and $m_1 v_1 + \ldots + m_h v_h + m_{h+1} \varepsilon v_{h+1} + .. + m_n \varepsilon v_n + \varepsilon v_{n+1} + k v_{n+2} + v_{n+3} = (m_1, .., m_h, m_{h+1}\varepsilon, .., m_n \varepsilon, \varepsilon, k, 0) \in L$. So the lattice L contains the entire message about the plaintext.

In our algorithm, the shortest vector is v_{n+1} and $\|v_{n+1}\| = \sqrt{2}$. From section 2.2, we know that if we use SVP algorithm, the result will be v_{n+1} not the plaintext. If we use the CVP algorithm the result could be $zv_{n+1} \in L$ not the plaintext. Thus our algorithm is secure against the Lattice attack.

V. Conclusion

In this paper, we proposed probabilistic encryption way that can adapt into many kinds of knapsack cryptosystems. All the key generation method can be used directly. We just need to change the encryption form, corresponding change the decryption method. Section 4 is just an example of how to change the decryption method.

The little change of the encryption way will cause a great change in the properties of the knapsack system. This encryption way has many good performances: it has a very large information rate; the brute force attack is computationally infeasible; it can protect from Shamir secret recovery attack, low-density subset-sum attack and Lattice attack; what's more, the probabilistic encryption way is secure against statistical analysis attack.

VI. Acknowledgment

This research was supported by the twelfth five years project on crypto development fund of China (MMJJ201102016).

VII. References

[1] Wang, B. and Y. Hu, Quadratic compact knapsack public-key cryptosystem. Computers & mathematics with applications,. 59(1): 194—206 (2010).

[2] Diffie W, Hellman M. New directions in cryptography. IEEE Transactions on Information Theory 22(6): 644-654 (1976)

[3] Merkle, R. and M. Hellman, Hiding information and signatures in trapdoor knapsacks. IEEE Transactions on Information Theory, 24(5): 525—530 (1978)

[4] Chor B, Rivest R L. A knapsack-type public key cryptosystem based on arithmetic in finite fields. IEEE Transactions on Information Theory, 34(5): 901-909 (1988).

[5] Okamoto T, Tanaka K, Uchiyama S. Quantum public-key cryptosystems. Advances in Cryptology—CRYPTO 2000, pp.147-165, Springer Berlin Heidelberg (2000)

[6] J.C. Lagarias and A.M. Odlyzko, Solving low-density subset sum problems, Journal of the Association for Computing Machinery 32(1) 229–246 (1985).

[7] Nguyễn P Q, Stern J. Adapting density attacks to low-weight knapsacks[M]. Advances in Cryptology, pp.41-58. Springer Berlin Heidelberg (2005)

[8] Coster M J, Joux A, LaMacchia B A, et al. Improved low-density subset sum algorithms. Computational Complexity, 2(2): 111-128 (1992).

[9] Gallo G, Hammer P L, Simeone B. Quadratic knapsack problems[M]. Combinatorial Optimization. pp.132-149. Springer Berlin Heidelberg (1980)

[10] Skiena S S. Who is interested in algorithms and why? lessons from the Stony Brook algorithms repository[J]. ACM SIGACT News, 30(3): 65-74(1999)

[11] Katayanagi K, Murakami Y, Kasahara M. A new product-sum public-key cryptosystem using message extension[J]. IEICE transaction on Fundamentals of Electronics, 84(10): 2482-2487 (2001).

[12] Lenstra A K, Lenstra H W, Lovász L. Factoring polynomials with rational coefficients[J]. Mathematische Annalen, 261(4): 515-534 (1982)

[13] Shor, P.W., Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Journal on Computing, 26(5): 1484--1509 (1997).

[14] Lvxin, Fengdengguo, quantum algorithm analysis of knapsack problem. Journal of Beijing university of aeronautics and astronautics, 30(11):1088-1091(in Chinese) (2004).

[15] Kunihiro N. New definition of density on knapsack cryptosystems[M]. Progress in Cryptology-AFRICACRYPT 2008 pp.156-173. Springer Berlin Heidelberg (2008)

[16] Murakami Y, Katayanagi K, Kasahara M. A new class of cryptosystems based on Chinese remainder theorem. Information Theory and Its Applications, (2008).

[17] B. Wang, Y. Hu, Knapsack-type public-key cryptosystem with high density. Journal of Electronics & Information Technology, 28 (12) 2390-2393.(in Chinese) (2006).

[18] Wang B, Wu Q, Hu Y. A knapsack-based probabilistic encryption scheme. Information Sciences, 177(19): 3981-3994(2007).