Cryptanaylsis of Knapsack Cryptosystem

Rajendra Kumar

April 2017

1 Introduction

Subset sum problem is a NP-complete problem[2]. Based on this problem knapsack cryptosystem was given by Merkle and Hellman[4]. In 1982 shamir[6] found the first attack on these cryptosystem by using the LLL algorithm. This report is on the Cryptanlysis of knapsack cryptosystem by Frieze[1]. Second section covers the fundamental problem and section 3 covers the details about the cryptosystem. In section 4 complete analysis of attack on knapsack cryptosystem is covered.

2 Subset Sum Problem

Definition 2.1 *Given a set $T = \{a_1, \ldots, a_n\}$ and $S \in \mathbb{Z}_M$. Find $\mathbf{x} \in \{0,1\}^n$ such that*

$$S = \sum_{i=1}^{n} x_i a_i \mod M$$

In general, solving subset sum problem is NP-Complete.

2.1 Easy problem

Definition 2.2 *A sequence $a_1, \ldots, a_n$ is super-increasing if*

$$a_i > \sum_{j=1}^{i-1} a_j, n \geq i > 1,$$

It is easy to see that there is a linear time greedy algorithm for solving the subset sum problem of super-increasing sequence.

3 Knapsack Cryptosystem

We know that general subset sum problem is hard to solve and subset sum problem of super-increasing sequence is easy to solve. From these two problem, we want to design a cryptosystem such that subset sum problem for receiver is easy to solve but for eavesdropper the subset sum problem should be hard to solve. By this approach Merkle and Hellman designed the knapsack cryptosystem in 1978[4].

3.1 Description of Cryptosystem

Private Key- Consist of $\{a'_1, \dots a'_n\}$ super-increasing sequence of n numbers, a prime number M such that $M > \sum_{i=1}^{n} a'_i$ and a multiplier w randomly choosen from $\mathbb{Z}^*_M$.
Generate $\{a_1, a_2, \dots, a_n\}$ where $a_i = wa'_i \mod M$.
Public Key- Consist of $\{a_1, a_2, \dots, a_n\}$ sequence of n numbers and prime number M.
Encryption- To encrypt a message $m \in 0, 1^n$. Generate cipher text

$$C = \sum_{i=1}^{n} m_i a_i \mod M$$

Decryption- To decrypt the cipher text C. We know that

$$w^{-1}C = \sum_{i=1}^{n} w^{-1} a_i x_i \mod M$$

$$w^{-1}C = \sum_{i=1}^{n} a'_i x_i \mod M$$

We know that above knapsack problem is easy to solve. So Encryption and Decryption can be efficiently done but for eavesdropper to find the secret message is hard.

4 Cryptanylsis of Knapsack Cryptosystem

Frieze showed that if the a_i are uniformly random in $\{1, \dots, M\}$ and $M \geq 2^{n^2(\frac{1}{2}+\epsilon)}$ then we can efficiently solve the subset sum problem with very high probability over the choice of the a_i.
We are given a subset sum problem instance with sequence $\mathbf{a} = \{a_1, \dots, a_n\}$ and number C. We want to find the $\mathbf{x} \in \{0, 1\}^n$ such that

$$C = \sum_{i=1}^{n} x_i a_i$$

Without loss of generality, we assume that $C \geq (\sum_{i=1}^{n} a_i)/2$, if not then we will replace C by C' such that $C' = (\sum_{i=1}^{n} a_i) - C$ and in the end we will the flip the bits of the answer $\mathbf{x}$ which we will find.
Let $B = \lceil (n2^n)^{1/2} \rceil$ and we generate a Lattice using basis matrix

$$L = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \\ -Ba_1 & -Ba_2 & \dots & -Ba_n & BC \end{bmatrix}$$

By using LLL , we can find a vector of lattice within length $2^{n/2}$ factor of $\lambda_1(L)$(length of shortest

non-zero vector in lattice). By analysis we are going to show that with high probability, we will obtain vector of the form $k \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}$ where k is a non-zero integer.

We know that length of the vector $\begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}$ is less than or equal to $n^{1/2}$. From above basis matrix we can say that last coordinate of all lattice vector is divisible by B. If last coordinate is non-zero then vector has length at least $B > 2^{n/2}n^{1/2} \geq 2^{n/2}\lambda_1(L)$. Therefore by LLL, we will always get vector with final coordinate zero.

Now, consider an arbitrary non-zero lattice vector $\begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix}$ where $||\mathbf{z}|| < 2^{n/2}n^{1/2}$. We are going to assume that $\mathbf{z}$ is not an integer multiple of $\mathbf{x}$ and we want to bound the probability of this vector where $\begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix} = L \begin{bmatrix} \mathbf{z} \\ z_{n+1} \end{bmatrix}$

We can say that,

$$C|z_{n+1}| = |\sum_{i=1}^{n} a_i z_i| \leq ||\mathbf{z}|| \sum_{i=1}^{n} a_i$$

We already assumed that $C \geq (\sum_{i=1}^{n} a_i)/2$. By this we can say that $|z_{n+1}| \leq 2||\mathbf{z}||$. For a fix value of z_{n+1}, we can say that

$$\sum_{i=1}^{n} a_i z_i = z_{n+1} C = z_{n+1} \sum_{i=1}^{n} a_i x_i$$

Which also implies that $\sum_{i=1}^{n} a_i y_i = 0$ where $y_i = z_i - z_{n+1} x_i$. We assumed that $\mathbf{z}$ is not an integer multiple of $\mathbf{x}$ so, there exist some i such that $y_i \neq 0$. Without loss of generality we can assume that $i = 1$. Therefore, we must require that $a_1 = -(\sum_{i=2}^{n} a_i y_i)/y_1$.

Now we want to find the probability of $\begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix} \in L$ for fixes $\mathbf{z}$,z_{n+1} is

$$Pr[\sum_{i=1}^{n} a_i y_i = 0] = Pr[a_1 = -(\sum_{i=2}^{n} a_i y_i)/y_1] \leq \frac{1}{M}$$

Because the a_i are chosen uniformly from $\{1, \ldots, M\}$.

We know that $||\mathbf{z}|| < B$ and $|z_{n+1}| < 2||\mathbf{z}|| < 2B$. Now we want to put the bound on number of choices of $\mathbf{z}, z_{n+1}$ which satisfy the above given condition and the bound is

$$(2B+1)^n(4B+1) \leq (5B)^{n+1} \leq 2^{n^2(\frac{1}{2}+\mathcal{O}(1))}$$

Therefore, if we take $M = 2^{n^2(\frac{1}{2}+\epsilon)}$ for $\epsilon > 0$, then the probability that there exist any $\begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix} \in L$ satisying the above condition is at most $2^{-\Omega(n^2)}$ which is extremely samall. Hence with very high probability LLL algorithm will give a vector of form $k \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix}$ and by this we can find the message $\mathbf{x}$.

References

[1] A M Frieze. On the lagarias-odlyzko algorithm for the subset sum problem. *SIAM J. Comput.*, 15(2):536–539, May 1986.

[2] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990.

[3] A. K. Lenstra, H. W. Lenstra, and L. Lovasz. Factoring polynomials with rational coefficients. *MATH. ANN*, 261:515–534, 1982.

[4] R. Merkle and M. Hellman. Hiding information and signatures in trapdoor knapsacks. *IEEE Trans. Inf. Theor.*, 24(5):525–530, September 2006.

[5] Chris Peikert. Lattices in cryptography 2013.

[6] Ad Shamir and N Diffie. A polynomial-time algorithm for breaking the basic merkle-hellman cryptosystem. In *In Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science*, pages 145–152. IEEE, 1982.