LATTICE BASED TOOLS IN CRYPTANALYSIS FOR PUBLIC KEY CRYPTOGRAPHY

R. Santosh Kumar[1], C. Narasimham[2] and S. Pallam Setty[3]

[1]Department of Information Technology ,MVGR College of Engg., Vizianagaram, India.
santu_hcunitk@yahoo.co.in
[2]Department of Information Technology,VR Siddhartha Engineering College, Vijayawada-7, India.
narasmham_c@yahoo.com
[3]Dept. of Computer Science & Systems Engineering, Andhra University, Vishakapatnam, India
drspsetty@yahoo.com

Abstract.

Lattice reduction is a powerful concept for solving diverse problems involving point lattices. Lattice reduction has been successfully utilizing in Number Theory, Linear algebra and Cryptology. Not only the existence of lattice based cryptosystems of hard in nature, but also has vulnerabilities by lattice reduction techniques. In this survey paper, we are focusing on point lattices and then describing an introduction to the theoretical and practical aspects of lattice reduction. Finally, we describe the applications of lattice reduction in Number theory, Linear algebra.

Keywords

Lattices, Lattice Reduction, RSA, Coppersmith, Subset Sum, Simultaneous Diophantine, Merkle-Hellman.

1 Introduction

Lattices are periodic arrangements of discrete points. Apart from their wide-spread use in pure mathematics, lattices have found applications in numerous other fields as diverse as cryptography/cryptanalysis, the geometry of numbers, factorization of integer polynomials, subset sum and knapsack problems, integer relations and Diophantine approximations, coding theory. In this paper, we survey the main tools which can be used to the verify vulnerabilities of different cryptosystems.

Lattice reduction is concerned with finding improved representations of a given lattice using algorithms like LLL (Lenstra, Lenstra, Lov´asz) reduction .There are some versions for lattice reduction, but people are using the LLL algorithm for theoretical and practical purposes. It is a polynomial time algorithm and the vectors are nearly orthogonal. In section II, we briefly discuss the complexity issues of LLL algorithm and its properties. In section III, we discuss the subset sum problem and how lattice reduction has been used to get a solution in some instances. This technique, in turn can be applied to break knapsack cryptosystems like Merkle-Hellman

DOI : 10.5121/ijnsa.2012.4212

knapsack cryptosystem. In section IV, we discuss Univaraite polynomial congruence problem and how lattice reduction was used to get a solution. This technique, in turn can be applied to check vulnerabilities of RSA cryptosystem. In section V, we discuss simultaneous Diophantine approximation problem and vulnerabilities of knapsack cryptosystem.

2. Terminology

2.1 Lattices

A lattice is a discrete subgroup of $\mathbb{R}^n$. Equivalently, given $m \leq n$ linearly independent vectors $b_1, b_2, b_3, \ldots, b_m \in \mathbb{R}^n$, the set $\mathcal{L} = \mathcal{L}(b_1, b_2, b_3, \cdots, b_m) = \{\sum_{i=1}^{m} \alpha_i b_i \,|\alpha_i \in \mathbb{Z}\}$, is a lattice. The $b_i$are called basis vectors of $\mathcal{L}$ and $\mathcal{B} = \{b_1, b_2, \cdots, b_m\}$ is called a lattice basis for $\mathcal{L}$. Thus, the lattice generated by a basis $\mathcal{B}$ is the set of all integer linear combinations of the basis vectors in $\mathcal{B}$. The determinant of a lattice, denoted by $vol(\mathcal{L})$ is the square root of the gramian determinant$det_{1\leq i,j\leq m}\langle b_i, b_j\rangle$, which is independent of particular choice of basis. A general treatment of this topic see[1].

2.2 Lattice reduction

Lattice reduction techniques have a long tradition in mathematics in the field of number theory. The goal of lattice basis reduction is to find, for a given lattice, a basis matrix with favorable properties. Usually, such a basis consists of vectors that are short and therefore this basis is called reduced. Unless stated otherwise, the term "short" is to be interpreted in the usual Euclidean sense. There are several definitions of lattice reduction with corresponding reduction criteria, such as Minkowski reduction, Hermite-Korkine-Zolotareff reduction, Gauss reduction, Lenstra-Lenstra-Lov´asz (LLL) reduction, Seysen reduction. The corresponding lattice reduction algorithms yield reduced bases with shorter basis vectors and improved orthogonality; they provide a tradeoff between the quality of the reduced basis and the computational effort required for finding it. Here we consider the LLL reduced, because there is a polynomial time algorithm exists and vectors are near orthogonal and the first vector solves the approximate SVP problem. For good survey on lattice reduction algorithms refers [4].

2.3 LLL reduced

The following LLL reduced version given by Lenstra, Lenstra, Lovasz[1],[2],[3].

LLL reduced: A basis $b_1, b_2, b_3, \cdots, b_n$ of a lattice $\mathcal{L}$ is said to be Lovasz-reduced or LLL-reduced if

$|\mu_{i,j}| \leq \frac{1}{2}$ for $1 \leq j < i \leq n$

$\left|b_i^* + \mu_{i,i-1} b_{i-1}^*\right|^2 \geq \frac{3}{4}|b_{i-1}^*|^2$ for $1 < i \leq n$. where the b_i^* and $\mu_{i,j}$ are defined by the Gram-Schimdt orthogonalization process acting on the b_i. Above in place of ¾ one can replace any quantity $\frac{1}{4} < \delta < 1$.

2.4 LLL Algorithm

The Lenstra –Lenstra -Lov´asz (LLL) algorithm [1][2][3] is an iterative algorithm that transforms a given
lattice basis into an LLL-reduced one. Since the definition of LLL-reduced uses Gram-Schmidt process, the LLL algorithm performs the Gram-Schimdt method as subroutine.

LLL Algorithm with Euclidean norm:

Input**:** $b_1, b_2, b_3, \cdots, b_n \in \mathbb{Z}^m$
Output: LLL reduced basis $b_1, b_2, b_3, \cdots, b_n$
1: Compute the Gram-Schimdt basis $b_1^*, b_2^*, \cdots, b_n^*$ and coefficients $\mu_{i,j}$ for $1 \leq j < i < n$.
2: Compute $B_i = \langle b_i^*, b_i^* \rangle = \|b_i^*\|^2$ for $1 \leq i \leq n$
3: k=2
4: while $k \leq n$ do
5: for $j = k - 1$ downto 1 do
6: let $q_j = \mu_{k,j}$ and set $b_k = b_k - q_j b_j$
7: update the values $\mu_{k,j}$ for $1 \leq j < k$ and B_k
8: end for
9: if $B_k \geq (\frac{3}{4} - \mu_{k,k-1}^2) B_{k-1}$ then
10: $k = k + 1$
11: else
12: Swap b_k with b_{k+1}
13: Update the values $b_k^*, b_{k-1}^*, B_k, B_{k-1}, \mu_{k-1,j}$ and $\mu_{k,j}$ for $1 \leq j < k$ and $\mu_{i,k}$ and $\mu_{i,k-1}$
for $k < i \leq n$.
14: $k = \min\{2, k - 1\}$
15: end if
16: end while

It can be proven that the LLL algorithm terminates a finite number of iterations. Let $\mathcal{L} \subset \mathbb{Z}^n$ be a lattice with basis$\{b_1, b_2, b_3, \cdots, b_m\}$ and $C \in \mathbb{R}$, $C \geq 2$ be such that $\|b_i\| \leq \sqrt{C}$ for $i = 1,2, \cdots, n$. Then the number of arithmetic operations needed for the algorithm $O(n^4 \log C)$ on integers of size $O(n \log C)$ bits.

2.5 Properties of LLL algorithm:

Let $b_1, b_2, b_3, \cdots, b_m$ be an LLL reduced basis for a lattice $\mathcal{L} \subset \mathbb{R}^m$. Then

1) $d(\mathcal{L}) \leq \prod_{i=1}^{n} |b_i| \leq 2^{\frac{n(n-1)}{4}} d(\mathcal{L})$,
2) $|b_j| \leq 2^{\frac{i-1}{2}} |b_i^*|$, if $1 \leq j \leq i \leq n$,
3) $|b_1| \leq 2^{\frac{n-1}{4}} d(\mathcal{L})^{\frac{1}{n}}$,
4) For every $x \in \mathcal{L}$ with $x \neq 0$ we have $|b_1| \leq 2^{\frac{n-1}{2}} |x|$.

In this paper, we use the property 4 frequently.

3. Solving subset sum problem of low density:

Let $\{a_1, a_2, a_3, \dots, a_n\}$ be distinct positive integers. The subset sum problem is, given an integer s obtained as a sum of elements a_i, to find $x_i \in \{0,1\}$ for $i = 1,2, \cdots n$ such that $\sum a_i x_i = s$. The density of *S* is defined to be $d = \frac{n}{\max\{\log a_i | 1 \le i \le n\}}$. The subset sum problem is $\mathcal{NP}$-complete.

3.1 LLL algorithm solution

Using LLL algorithm one can find a particular short vector in a lattice[4]. Since the reduced basis produced by LLL algorithm includes a vector of length which is guaranteed to be within a factor of $2^{\frac{n-1}{2}}$ of the shortest non-zero vector of the lattice. In practice, however, the LLL algorithm usually finds a vector which is much shorter than what is guaranteed. So the LLL algorithm can be expected to find the short vector which yields a solution to the subset sum problem provided that this vector is shorter than most of the non zero vectors in the lattice.

3.2 Justification

Consider the matrix $(n+1) * (n+2)$ matrix $B = \begin{bmatrix} 2 & 0 & 0 & & 0 & ma_1 & 0 \\ 0 & 2 & 0 & \cdots & 0 & ma_2 & 0 \\ 0 & 0 & 2 & & 0 & ma_3 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & & 2 & ma_n & 0 \\ 1 & 1 & 1 & \cdots & 1 & ms & 1 \end{bmatrix}$

Let the rows of the matrix B be $b_1, b_2, b_3, \dots, b_n, b_{n+1}$ and L be the lattice generated by these vectors. If $x_1, x_2, x_3, \dots, x_n$ is a solution to the subset sum problem, then we have

$$y = \sum_{i=1}^{n} x_i\, b_i - b_{n+1}$$

$$= (x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_n b_n - b_{n+1})$$

$$= (2x_1 - 1, 2x_2 - 1, \dots, 2x_n - 1, m(a_1x_1 + a_2x_2 + \dots + a_nx_n - s), 1)$$

Since $(x_1, x_2, x_3, \dots, x_n)$ is a solution and each x_i $(1 \le i \le n)$ is either 0 or 1, we have $y_i \in \{-1,1\}$ and $y_{n+1} = 0$. Since $\|y\| = \sqrt{y_1^2 + \dots + y_{n+1}^2 + y_{n+2}^2}$, the vector *y* is a vector of short length in *L*. If the density of the knapsack set is small, i.e the a_i are large, then most vectors in *L* will have relatively large lengths, and hence *y* may be unique shortest non zero vector in *L*. If this is indeed the case then there is a good possibility of the algorithm finding a basis which includes this vector. Above algorithm is not guaranteed to succeed. Assuming that the LLL algorithm always produces a basis which includes the shortest non zero lattice vector, then algorithm succeeds with high probability if the density of the knapsack set is less than 0.9408.

3.3 Application

This is most powerful general attack known on knapsack encryption schemes[5]. It is typically successful if the density of the knapsack set is less than 0.9408. This is significant because the

density of a Merkle-Hellman knapsack[6] set much be less than 1, since otherwise there will be many subsets of the knapsack set with the same sum, in which case some cipher texts will not be uniquely decipherable. Also, since each iteration in the multiple-iterated scheme lowers the density, this attack will succeed if the knapsack set has been iterated a sufficient number of times. Similar techniques have since been used to break most knapsacks schemes that have been proposed, including the multiple-iterated Merkle-Hellman scheme.

4. Solving modular equations

It is easy to compute the integer roots of a polynomial in a single variable over the integers. But the related problem of solving modular equations can be hard. We have different tools to solve $f(x) = 0$. But one cannot solve $f(x) = 0\ (mod\ n)$ efficiently. The solution for the above equation was proposed by Coppersmith in the year 1997[7]. Here we present simple version of Howgrave-Graham[8].

Let N be an integer and $f \in \mathbb{Z}[x]$ be a monic polynomial of degree d. Set $X = N^{\frac{1}{d}-\varepsilon}$ for some $\varepsilon \geq 0$. Then given $\langle N, f \rangle$, one can efficiently find all integers $|x_0| < X$ satisfying $f(x_0) = 0(mod\ N)$ using the LLL algorithm. This fact claims the existence of an algorithm which can efficiently find all roots of f modulo N that are less than $X = N^{\frac{1}{d}}$. As X gets smaller, the algorithm's runtime decreases. This theorem's strength is the ability to find out all small roots of polynomials modulo a composite N. The idea is simply reducing the root finding problem in modular equations to the case of root finding equations over the integers. Thus one has to construct from the polynomial $f_b(x)$ with the root $x_0 \leq X$ modulo b a polynomial of $f(x)$ which has the same root x_0 by applying standard root finding algorithms to $f(x)$. But how can be transform $f_b(x)$ into $f(x)$?. This transform is exactly the core of the Coppersmith's method. He defines the matrix which has the elements of the form $g_{i,j}(x) = N^{m-i}x^j f_b^i(x)$ for $i = 1, \cdots m$ and some choice of j and it has a root $x_0 mod\ b^m$. Then every integer linear combination $f(x) = \sum_{i,j} g_{i,j}(x),\ a_{i,j} \in \mathbb{Z}$ of polynomials in G also has the root $x_0 mod\ b^m$. Our goal is to find among these linear combinations one which has the root x_0 not just modulo b^m but also over the integers. For this one can choose coefficients of $f(x)$ satisfies the relation $f(x_0) < b^m$. This is where the lattice reduction algorithm such as LLL comes into the picture. The first vector of a reduced basis satisfies the above inequality.

4.1 Application 1: Attacking RSA with small e by knowing parts of the message:

Suppose that $m = M + x$ for some known part M of the message and some unknown part $x \leq N^{\frac{1}{e}}$. Now one can recover m from above scenario. This situation occurs in the case of stereotyped messages. Let (N, e) be a public key in RSA public key cryptosystem[7]. Furthermore, let $C = (M + x_0)^e mod\ N$ be an RSA encrypted message with known M and unknown x_0,where $|x_0| \leq N^{\frac{1}{e}}$. Then one can find x_0 in time polynomial in $\log N$ and e. The above fact is direct application of Coppersmith's method.

We can extend the above attack to hold in the case where the unknown part x is somewhere in the middle of the message i.e $m = M + x2^k + M'$ when x begins at the $(k+1)^{st}$

least significant bit. Here we need small modification to get monic polynomial from the equation $f(x) = (M + x2^k + M')^e - c$ by multiplying with 2^{-ke}.

4.2 Application 2: Repeated message and short pad attack

Consider the situation when Bob sends two messages to Alice that only differ by a small amount. Also
assume that sender is using a exponent 3. In this case $M^3 = c_1(mod\ N)$ and $(M + x)^3 = c_2(mod\ N)$.
One can eliminate the M from above two equations by using resultants, and is left with the equation
$x^9 + 3(c_1 - c_2)x^6 + 3(c_2^2 + 7c_1c_2 + c_1^2)x^3 + (c_1 - c_2)^3 = 0(mod\ N)$, so one may discover the padding
as long as $|x| \leq N^{\frac{1}{9}}$. It is not obvious that recovering M from the knowledge of x, but this is true due to
clever trick of Franklin and Reiter[9]. To explain this in our case let m be a polynomial indeterminate and
calculate gcd $(m^3 - c_1, (m + x)^3 - c_2)$ using the Euclidean algorithm. It can be shown that the result of
this gcd will be the linear polynomial $m - M$, and hence we discover m.

5.Simultaneous Diophantine Approximation

Simultaneous Diophantine approximation is concerned with approximating a vector $\left(\frac{q_1}{q}, \frac{q_2}{q}, \cdots, \frac{q_n}{q}\right)$ of rational numbers by a vector of $\left(\frac{p_1}{q}, \frac{p_2}{q}, \cdots, \frac{p_n}{q}\right)$ of rational numbers with a smaller denominator p. Algorithms for finding simultaneous Diophantine approximation have been used to break some knapsack public key cryptosystems. The vector $\left(\frac{p_1}{q}, \frac{p_2}{q}, \cdots, \frac{p_n}{q}\right)$ of rational numbers is said to be a simultaneous Diophantine approximation of δ-quality to the vector $\left(\frac{q_1}{q}, \frac{q_2}{q}, \cdots, \frac{q_n}{q}\right)$ of rational numbers if $p < q$ and $\left|p\frac{q_i}{q} - p_i\right| \leq q^{-\delta}$ for $i = 1,2,\cdots,n$. One can reduce the problem of finding a δ-quality simultaneous Diophantine approximation to the problem of finding a short vector in a lattice [2]. The latter problem can be solved using LLL algorithm. Consider the (n+1) dimensional matrix $A_{i,j}$ as
$=\lambda q\ if\ i = j$ and $1 \leq i \leq n$,
=0 if i≠ j and $i \leq i \leq n$,
=-λq_jif $i = n + 1$ and j≠ $n + 1$,
=1 if $i = n + 1$and j=$n + 1$ where $\lambda \approx q^{\delta}$.

5.1 Justification:

Apply LLL algorithm to above matrix and let the rows of the matrix A be denoted by $(b_1, b_2, \cdots, b_n, b_{n+1})$. Suppose that $\left(\frac{q_1}{q}, \frac{q_2}{q}, \cdots, \frac{q_n}{q}\right)$ has a δquality approximation

$\left(\frac{p_1}{q}, \frac{p_2}{q}, \cdots, \frac{p_n}{q}\right)$. Then $x = p_1 b_1 + p_2 b_2 + \cdots + p_n b_n + p_{n+1} b_{n+1}$=$(\lambda(p_1 q - p q_1), \lambda(p_2 q - p q_2), \cdots, \lambda(p_n q - p q_n), p)$is in L and has length less than approximately $\sqrt{n+1}q$. Thus x is short compared to the original basis vectors, which are of length roughly $q^{1+\delta}$. Also, if $v = (v_1, v_2, \cdots, v_n, v_{n+1})$ is a vector in L of length less than q, then the vector $\left(\frac{p_1}{q}, \frac{p_2}{q}, \cdots, \frac{p_n}{q}\right)$ defined as above is a δ quality approximation.

5.2 Application

Given the public knapsack set, this technique finds a pair of integers U', M' such that $\frac{U'}{M'}$ is close to $\frac{U}{M}$ where U and M are part of the private key of the Merkle-Hellman Cryptosystem and $U = W^{-1} mod\, M$ and such that the integers $b_i' = U' a_i mod\, M, 1 \leq i \leq n$ form a super increasing sequence. This sequence can then used by an adversary to decrypt messages [2].

6. Conclusions:

In this survey paper, we have discussed some Cryptanalytic attacks using some tricky lattice techniques. First one, we solved subset sum problem of low density. Then we observe vulnerabilities of Merkle-Hellman knapsack cryptosystem which is based on subset sum problem. Second one, we solved univaraite modular polynomial equations. Using this we check the pitfalls of RSA function in two cases. Finally we discuss the problem of Simultaneous Diophantine Approximation problem. Again we observe vulnerabilities of Merkle-Hellman Cryptosystem. All are implemented in NTL number theory [12] library maintaining by victor shoup.

References:

1. H. Cohen, A Course in computational Algebraic Number Theory. Springer-Verlag, second edition, 1995.
2. A. J. Menezes, P.C.van Oorschot, and S.A.Vanstone. Hand book of Applied Cryptography. CRC Press, 1997.
3. A.K. Lenstra, H.W. Lenstra Jr., and L.Lovasz. Factoring polynomials with rational coefficients. Mathematische Annalen, volume 261(4): pages 515-534, 1982.
4. P. Schnorr and M. Euchner, Lattice basis reduction: Improved practical algorithms and solving subset sum. problem .Math.Prog., 66: 181- 199,1994
5. A. Shamir. A polynomial-time algorithm for breaking the basic Merkle-Hellman cryptosystem. IEEE . Transactions on Information Theory, 1984
6. R .C. Merkle and M.E. Hellman. Hiding information and signatures in trapdoor knapsacks. IEEE Transactions on Information Theory, IT-24(5):525-530, September 1978.
7. R.L.Rivest, A.Shamir and L.Adleman. A method for obtaining digital signatures and public key cryptosystems. Commun. Of the ACM, 21: 120-126, 1978.

8. D.Coppersmith. Finding a small root of a univaraite modular equation. Advances in Cryptology Proceedings of EUROCRYPT'96, VOLUME 1070 of Lecture Notes in Computer Science, pages 155-165. Springer Verlag, 1996.
9. N.A. Howgrave - Graham. Finding small solutions of univaraite modular equations revisited. In Cryptography and Coding, volume 1355 of LNCS, pages131-142. Springer-Verlag, 1997.
10. Don Coppersmith, Matthew Franklin, Jacques Patarin, Michael Reitert. Low Exponenet RSA with related messages.
11. R.Kannan. Algorithmic geometry of numbers. Annual Review of Computer Science, (231-267), 1987.
12. Victor Shoup. NTL: A library for doing number theory. Website: http://www.shoup.net/ntl/