Survey: Lattice Reduction Attacks on RSA

David Wong
supervised by Guilhem Castagnos

University of Bordeaux, March 2015

Abstract

RSA, carrying the names of **Ron Rivest**, **Adi Shamir** and **Leonard Adleman**, is one of the first practicable **public-key** cryptosystem. The algorithm was publicly described for the first time in **1977** and has since been the most used cryptosystem when it comes to asymmetric problems. For now more than **30 years**, Cryptanalysts and Researchers have looked for ways to **attack RSA**.
One branch of cryptanalysis on RSA is take an interest in a **relaxed model** of RSA. A model where we know part of the message, or we know an approximation of the primes, or the private exponent is too small... In these sort of problems, **lattice reduction techniques** have proved to be very relevant. **Coppersmith** opened the way with his constructive theorem on how to find small roots of **univariate** polynomials using reductions of lattices. **Boneh** and **Durfee** recently followed with a method on how to find small roots of **bivariate** polynomials using Coppersmith's heuristics on multivariate polynomials. In this survey we will see how each algorithm work and how they were respectively made simpler by **Howgrave-Graham** and the duo **Herrmann** and **May**.

Keywords : RSA, lattice, LLL, Coppersmith, Howgrave-Graham, Boneh-Durfee, Herrmann-May.

1 Introduction

In 1995, **Coppersmith** released a paper on how to attack RSA using **Lattices** and **Lattice reduction techniques** such as **LLL**. A few years later, **Howgrave-Graham** revisited Coppersmith's algorithm and made it easier to understand and apply. His work was implemented for various problems from revealing part of a message if most of the message was known, to breaking RSA if a good enough approximation of one of the prime was known. Attacks based on Lattice reduction techniques caught up and several researches were done on the subject. In 1990, **Wiener** had found that you

could successfully break RSA if the private exponent was too small ($d < N^{1/4}$). In 2000, **Boneh** and **Durfee** improved that bound ($d < N^{0.292}$) using lattices and LLL in a Coppersmith-like attack. Their work was later simplified by **Herrmann** and **May**.
In the 2nd and 3rd sections of this survey I will briefly explain how RSA and Lattice work. In section 4 we will see in what **model** the attacks are taking place and see **Håstad's Broadcast Attack** as an introduction to Coppersmith. Section 5 will be an overview of the Coppersmith algorithm revisited by Howgrave-Graham. Section 6 will be an overview of the Boneh and Durfee algorithm revisited by Herrmann and May. Finally the **implementations** of both attack will be added as an appendix.

2 RSA

Let's quickly recall **what is** and **how** RSA works :
RSA is an **asymmetric cryptosystem**. A generator algorithm derives two kinds of keys : a **public key** and a **private key**, both can be used either to encrypt or decrypt thanks to the asymmetric property of RSA to allow us to use the system as an **encryption** system or as a **signature** system.

2.1 Generation

To use RSA for **encryption** we need a public key to encrypt and a private key to decrypt. We first generate the **public key** as follow :

We generate **two primes** p and q. For security issues they should be around the **same size**. Those primes are the **core elements** of RSA. Knowing one of those allows us to compute the private key, thus allowing us to break the system. They can also be used to speed up calculations using the Chinese Remainder Theorem.
Knowing p and q we can then compute the **modulus** $N = p \times q$ which will be part of the public key as well as the private key. And you will see why.

Now comes the interesting part, we need to find a **public exponent** which will be used for **encryption**. For computation purpose a **Fermat prime** $(2^m + 1)$ is often used as public exponent (it makes things faster in **binary exponentiation**). In the case of a **signature scheme**, we would want the speed up to occur for the **private exponent** so we would use such a number as a private exponent and we would reverse the following steps.
Anyway, any kind of exponent can theoretically be chosen, as long as it is coprime with $\varphi(N)$ (The **Euler's Totient function**).

$$e \leftarrow \mathbb{Z}^*_{\varphi(N)}$$

if e is **coprime** with $\varphi(N)$ then it is part of the multiplicative group $(\mathbb{Z}/\varphi(N)\mathbb{Z})^*$ and thus **invertible** in $\mathbb{Z}/\varphi(N)\mathbb{Z}$.

Now it is pretty easy to find the private exponent d by inverting our public exponent e.

All of this is possible because we can easily compute $\varphi(N)$:

$$\varphi(N) = (p-1) \times (q-1)$$

And here resides the **security** of RSA. Imagine for a moment that we could easily factor N into p and q, then we would be able to invert the public exponent e. That's why we say that **the security of RSA is reduced to the Factorization Problem**.

Now let the **private key** be **(N, d)** with the addition of (p, q) if we need to speed up calculations. And let the **public key** be **(N, e)**.

2.2 Encryption/Decryption

To **encrypt** a message m, with $m < N$ we just do :

$$c = m^e \pmod{N}$$

And to **decrypt** :

$$m = c^d \pmod{N}$$

This works because the decryption step gives us :

$$c^d = (m^e)^d \pmod{N}$$

And e being d's inverse tells us that :

$$\begin{aligned} e &= d^{-1} \pmod{\varphi(N)} \\ \implies ed &= 1 \pmod{\varphi(N)} \\ \implies ed &= \varphi(N) + 1 \end{aligned}$$

Coupled with **Euler's Theorem** stating that if a and n are coprime then :

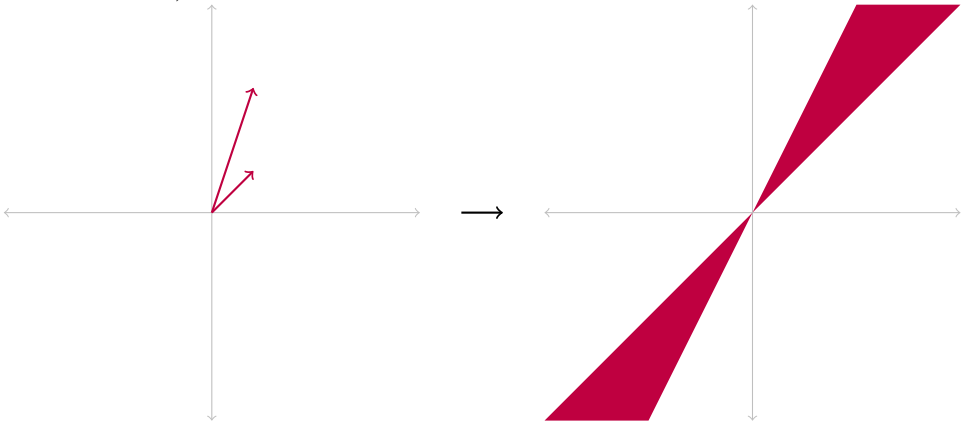
$$a^{\varphi(n)} = 1 \pmod{n}$$

Tells us that $m^{ed} = m \pmod{N}$

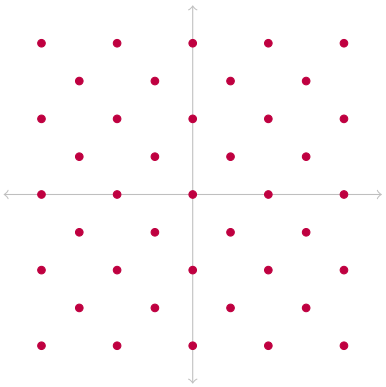
3 Lattice

3.1 Introduction

The attacks we will describe later both make use of the **Lenstra–Lenstra–Lovász lattice basis reduction algorithm**. Hence it is necessary for us to understand what is a lattice and why is this **LLL** algorithm so useful.
Think about Lattices like **Vector Spaces**. Imagine a simple vector space of two vectors. You can add them together, multiply them by scalars (let's say numbers of $\mathbb{R}$) and it spans a vector space.



Now imagine that our vector space's **scalars are the integers**, taken in $\mathbb{Z}$. The space spanned by the vectors is now made out of points. It's **discrete**. Meaning that for any point of this lattice there is a ball centered around that point of radius different from zero that contains only that point. Nothing else.



Lattice are interesting in cryptography because we seldom deal with real numbers and they bring us a lot of tools to deal with integers.
Just as vector spaces, lattices can also be described by different basis represented as **matrices**. Contrary to vector spaces, we generally represent the

vectors of the basis as **rows** in their corresponding matrices.

Last but not least, if $\{\tilde{b}_1 \ldots, \tilde{b}_w\}$ are the vectors of the Gram-Schmidt basis of a lattice L then we define the **determinant** of the lattice as such :

$$det(L) := \prod_{i=1}^{w} \|\tilde{b}_i\|$$

You will see that in the technique we present, to easily compute the determinant of a lattice we will make the lattice **full rank** (dimension = rank) and **triangular**. So that the determinant is computable by doing the products of the **diagonal terms** of the lattice basis.

3.2 LLL

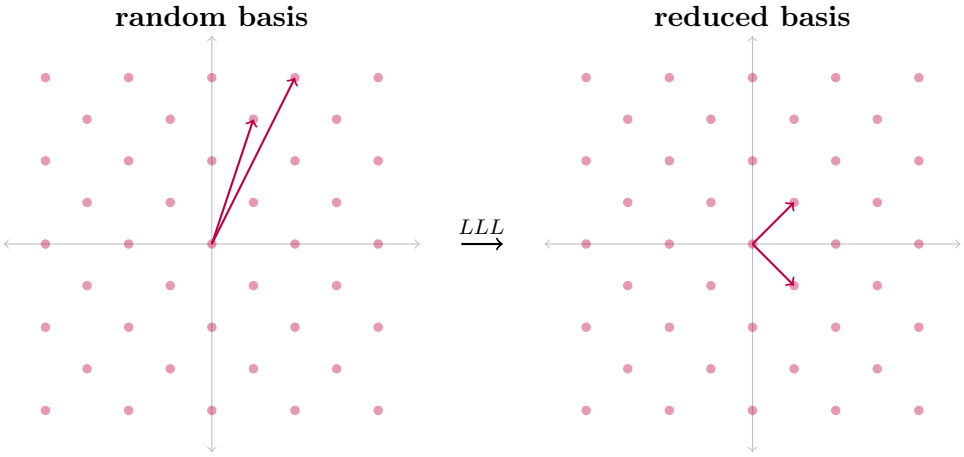
The **Lenstra–Lenstra–Lovász** *lattice basis reduction algorithm* is a step by step calculus that reduces a lattice basis in polynomial time. The lattice is left unchanged but the row vectors of its new basis are "**smaller**" according to some definitions :

Definition 1. *Let L be a lattice with a basis B. The δ-LLL algorithm applied on L's basis B produces a new basis of L : $B' = \{b_1, \ldots, b_n\}$ satisfying :*

$$\forall\, 1 \leq j < i \leq n \text{ we have } |\mu_{i,j}| \leq \frac{1}{2} \tag{1}$$

$$\forall\, 1 \leq i < n \text{ we have } \delta \cdot \|\tilde{b_i}\|^2 \leq \|\mu_{i+1,i} \cdot \tilde{b}_i + \tilde{b}_{i+1}\|^2 \tag{2}$$

$$\text{with } \mu_{i,j} = \frac{b_i \cdot \tilde{b}_j}{\tilde{b}_j \cdot \tilde{b}_j} \text{ and } \tilde{b}_1 = b_1 \text{ (Gram-Schmidt)}$$



We will not see dig into the interals of LLL here, see Chris Peikert's course[1] for a detailed view of the algorithm.

3.3 Wanted properties of LLL

LLL yields an approximation of the **Shortest Vector Problem**. This is useful for us because if we consider the row vectors of a lattice's basis as **coefficient vectors of polynomials**. We can find a **linear combination** of those polynomials that has **"particularly small" coefficients**. But let's not unveil too much too soon. Here are some relevant properties of a LLL reduced basis that we will need later :

Property 1. *Let L be a lattice of dimension n. In polynomial time, the LLL algorithm outputs reduced basis vectors v_i, for $1 \leq i \leq n$, satisfying :*

$$\|v_1\| \leq \|v_2\| \leq \ldots \leq \|v_i\| \leq 2^{\frac{n(n-1)}{4(n+1-i)}} \cdot det(L)^{\frac{1}{n+1-i}}$$

We can see that we can modify the bound on our vectors by modifying the dimension and the determinant of the lattice basis.

4 Relaxed models and small roots problem

Attacks on RSA falls into two categories : the attacks on the **implementation** or the **mathematical attacks**. Over the years the mathematical cryptanalysis on RSA have proven to be hard and thus the cryptosystem is still considered as secure nowadays (march 2015). But what a researcher could find interesting is to attack a **relaxed** model of the RSA problem. What if we knew "a part" of the message, or what if we knew "an approximation" of one of the prime, or what if the private exponent was "too small"...

Let's imagine for an instant that Alice used RSA to encrypt the same message to 3 different persons, all using the **same "small" public exponent** $e = 3$ as it's common to do. There is an attack, called **Håstad's Broadcast Attack**, that breaks this model.

$$\begin{aligned} c_1 &= m^3 \pmod{N_1} \\ c_2 &= m^3 \pmod{N_2} \\ c_3 &= m^3 \pmod{N_3} \end{aligned}$$

Here the trick is to use the **Chinese Remainder Theorem** to create an equation $\mod N_1 \cdot N_2 \cdot N_3$:

$$\begin{aligned} m^3 =& c_1 \cdot N_2N_3 \cdot [(N_2N_3)^{-1} \pmod{N_1}] \\ &+ c_2 \cdot N_1N_3 \cdot [(N_1N_3)^{-1} \pmod{N_2}] \\ &+ c_3 \cdot N_1N_2 \cdot [(N_1N_2)^{-1} \pmod{N_3}] \pmod{N_1N_2N_3} \end{aligned}$$

The method is similar to **Lagrange Interpolation**. For example let me quickly explain the first term, this m^e has to be equal to c_1 only when modulo N_1, so we can multiply the term c_1 by N_2 and N_3 so that it cancels out when modulo N_2 or N_3. But when it is modulo N_1 we don't want those terms, so we multiply our term by their inverse modulo N_1 as well. Easy no ? All the variables are known so calculating $m^e \pmod{N_1N_2N_3}$ is straight forward.
Let's notice that since $m < N_1, m < N_2, m < N_3$, we must have :

$$m \times m \times m = m^3 < N_1 \times N_2 \times N_3$$

So our $m^3 \pmod{N_1N_2N_3}$ is actually just m^3 over $\mathbb{Z}$.
To recover the message we just have to calculate the cubic root of that big value we just calculated.

Generalizing it is pretty easy and let's formulate **Håstad**'s findings :

Theorem 1. *If $c = m^e \pmod{N}$, then we can find m in time polynomial if $|m| < N^{1/e}$.*

That's the introduction to our "small root" problem. Now what about if $|m| > N^{1/e}$ but we know a part of the message m_0 :

$$c = (m_0 + x)^e \pmod{N}$$

Can we efficiently recover x ? That's the question Coppersmith is answering.

5 Coppersmith

This survey is no replacement for the original papers of Coppersmith[2] and Howgrave-Graham[3]. If you want to get a real understanding of those techniques I also advise you to read the survey from May[4].

5.1 Known modulus

That being said, let's dig into Coppersmith's use of LLL to crack RSA. We'll first see one of the problem it solves and build it from there.
Imagine that you know a part of the message : this is called the **Stereotyped Messages Attack**. For example you know that the Alice always sends her messages this way : "the password is : cupcake".

Let's say we know m_0 of the message $m = m_0 + x$. And of course we don't know x. We have **our problem** translated to the following equation :

$$f(x) = (m_0 + x)^e - c = 0 \pmod{N}$$

Well. **Coppersmith** says we can solve this in polynomial time if x and e are small enough :

Theorem 2. *Let N be an integer of unknown factorization, which has a divisor $b \geq N^\beta$, $0 < \beta \leq 1$. Let $f(x)$ be a univariate monic polynomial of degree δ and let $c \geq 1$.*
Then we can find in time $\mathcal{O}(c\delta^5 log^9(N))$ all solutions x_0 of the equation

$$f(x) = 0 \pmod{b} \quad with \quad |x_0| \leq c \cdot N^{\frac{\beta^2}{\delta}}$$

In our case that would mean that for $c = 1$ and $beta = 1$ we could find a solution of our previous equation if $|x_0| \leq N^{\frac{1}{e}}$. And here we find something very similar to Håstad's Broadcast Attack.
To find the roots of a polynomial **over a ring of integers modulo** N is a very **difficult** task, whereas we possess efficient tools to find roots of polynomials **over the integers** (Berlekamp–Zassenhaus, van Hoeij, Hensel lifting...). Hence Coppersmith's intuition to look for such a polynomial :

$f(x_0) = 0 \pmod{N}$ with $|x_0| < X$

$\downarrow$

$g(x_0) = 0$ over $\mathbb{Z}$

But how can we go from f to g here ? The theorem of **Howgrave-Graham** gives us a clue :

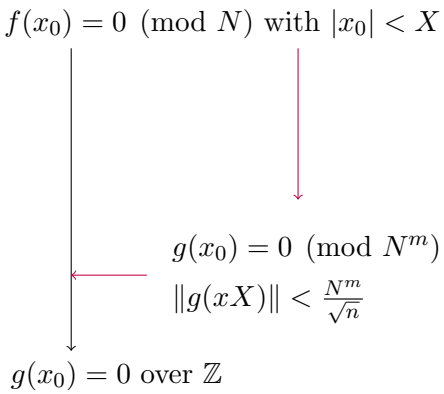
Theorem 3. *Let $g(x)$ be an univariate polynomial with n monomials. Further, let m be a positive integer. Suppose that*

$$g(x_0) = 0 \pmod{N^m} \quad where \quad |x_0| \leq X \tag{1}$$

$$\|g(xX)\| < \frac{N^m}{\sqrt{n}} \tag{2}$$

Then $g(x_0) = 0$ holds over the integers.

What Howgrave-Graham is saying is that we need to **find a polynomial** that shares the same root as our function f but modulo N^m and it has to have **"small" coefficients** so that its norm would be "small" as well.



Howgrave-Graham's idea is that we need to find this polynomial g by **combining** polynomials who also have x_0 as roots. The more polynomials we can play with, the better. We will see later that it is very easy for us to create polynomials f_i such that $f_i(x_0) = 0 \pmod{N^m}$. And that is the reason why we choose to find a polynomial over N^m and not over N.

The **LLL reduction** has two properties that are useful to us :
- It only does **integer linear operations** on the basis vectors
- The **shortest vector of the output basis is bound** (as seen in **Property 1**)

The first point allows us to combine them to build a function that still has x_0 as root modulo N^m :

$$g(x_0) = \sum_{i=1}^{n} a_i \cdot f_i(x_0) = 0 \pmod{N^m} \quad a_i \in \mathbb{Z}$$

The second point allows us to get Howgrave-Graham's second point ($\|g(xX)\| < \frac{b^m}{\sqrt{n}}$).

But first let's see how to **build the polynomials** f_i (we will call them $g_{i,j}$ and h_i) we will build our $g(x_0) = 0$ with. Note that δ is the degree of f :

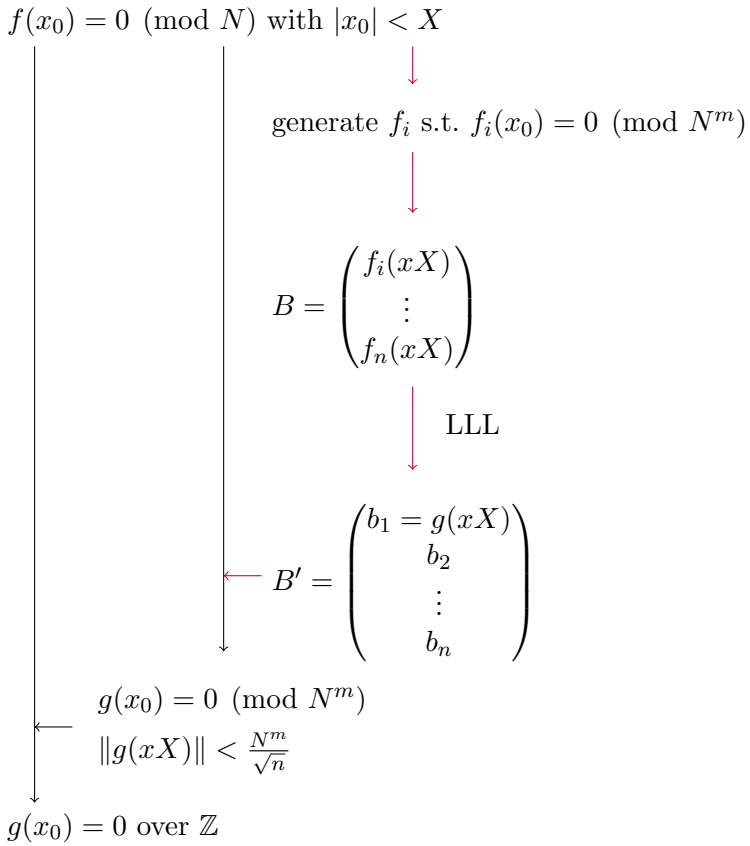
$$g_{i,j}(x) = x^j \cdot N^i \cdot f^{m-i}(x) \text{ for } i = 0, \dots, m-1, \;\; j = 0, \dots, \delta - 1$$
$$h_i(x) = x^i \cdot f^m(x) \text{ for } i = 0, \dots, t-1$$

Those polynomials achieve two things :
- they have the **same root** x_0 but modulo N^m
- each iteration introduce a new polynomial. That allows us to build a **triangular** lattice (so that the determinant is easier to calculate)

If you don't understand how they have the same root x_0 remember that since $f(x_0) = 0 \pmod N$ we know that $f(x_0) = k \cdot N$

Now we just have to create a lattice basis with $f_i(xX)$ as row vectors (because we want them to build a polynomial $g(xX)$ to test Howgrave-Graham second point).

Let's take a look at the **overview** again :



Now the shortest vector of B' (the LLL-reduced basis) should be the coefficient vector of $g(xX)$.

As I said earlier, the LLL reduction allows us to achieve an **upper bound** on this shortest vector :

$$\|b_1\| \leq 2^{\frac{n-1}{4}} \cdot det(L)^{\frac{1}{n}}$$

And recall **Howgrave-Graham Theorem**'s second point :

$$\|b1\| = \|g(xX)\| < \frac{N^m}{\sqrt{n}}$$

Now, to obtain Howgrave-Graham's second point on our g we have to manipulate $g_{i,j}(xX)$ and $h_i(xX)$ to obtain a small enough determinant. From the previous equations we **bound the determinant** :

$$det(L) < 2^{-\frac{n(n-1)}{4}} \cdot n^{-\frac{n}{2}} \cdot N^{n \cdot m}$$

The small terms can be considered as "error terms" to **simplify our bound** :

$$det(L) < N^{m \cdot n}$$

It is from these equations that Coppersmith bounded the value of x in his theorem. Now if we want to use this algorithm we will have to **tweak** m and t until we obtain the correct bounds. Note that the bound on the shortest vector of the reduced lattice basis is generous. That means that even if we don't correctly bound our determinant, we might find an answer.

5.2 Any modulus

Coppersmith method is actually more general : it also works for unknown modulus.
We will see how the **Factoring with High Bits Known** attack works to understand this part. Imagine the relaxed problem of RSA where we know an approximation $\tilde{p}$ of one of the prime p. The approximation is bounded as followed :

$$|\tilde{p} - p| < N^{\frac{1}{4}}$$

Now we have an equation with one unknown, modulo another unknown :

$$\tilde{p} = x_0 \pmod{p}$$

This gives us an equation $f(x) = \tilde{p} - x$ such that $f(x_0) = 0 \pmod{p}$. We can use that in the Coppersmith algorithm we have seen earlier. This is because Howgrave-Graham's theorem works for unknown modulus. Let's see this theorem again :

Theorem 4. *Let $g(x)$ be an univariate polynomial with n monomials. Further, let m be a positive integer. Suppose that*

$$g(x_0) = 0 \pmod{b^m} \quad \textit{where} \quad |x_0| \leq X \tag{1}$$

$$\|g(xX)\| < \frac{b^m}{\sqrt{n}} \tag{2}$$

Then $g(x_0) = 0$ holds over the integers.

We know we can build the f_i polynomials as we did before. And here instead of bounding $\|g(xX)\|$ with p^m we can bound it with $N^{\beta m}$ (since we have $p > N^\beta$ in Coppersmith Theorem). This allows us to formulate problems with unknown modulus.
And now obtain a bound on the determinant :

$$det(L) < N^{m \cdot n \cdot \beta}$$

5.3 How were the bounds calculated ?

And here's the **determinant** :

$$det(L) = N^{\frac{1}{2}\delta m(m+1)} X^{\frac{1}{2}n(n-1)}$$

5.4 Experiments

I used Sage 6.4 in a Virtualbox with 512Mo of RAM and 1 core from an Intel i7 @ 2.30GHz.
Here are the experiments for the **Stereotyped Message Attack** :

size of x_0	size of N	e	m	t	running time
100	512	3	3	0	0.02s
200	1024	3	3	0	0.05s

Here are the experiments for the **Factoring with High Bits Known Attack** :

size of $\|p - \tilde{p}\|$	size of N	m	t	running time
110	512	4	4	0.01s
200	1024	4	4	0.03s

6 Boneh-Durfee

6.1 Overview of the method

This survey is no replacement for the original papers of Boneh and Durfee[5] and Herrmann and May[6].
We've seen how Coppersmith found a way of using lattices and the LLL algorithm to find small roots to particular univariate polynomials. What about problems that have two unknowns ? Coppersmith gave an heuristic for finding roots of **bivariate polynomials** but left it at that. More recently, Boneh and Durfee have released papers on some RSA attacks that make use of the initial ideas of Coppersmith for finding small roots of bivariate polynomials.

Let's introduce the problem :
Boneh and Durfee are telling us we can, **most of the time** (they released a heuristic and not a theorem), successfully **factor N** if the **private exponent d is too small**. Precisely if $d < N^{0.292}$.

Recall how RSA works :

$$\begin{aligned} & e \cdot d = 1 \pmod{\varphi(N)} \\ \implies & e \cdot d = k \cdot \varphi(N) + 1 \\ \implies & k \cdot \varphi(N) + 1 = 0 \pmod{e} \\ \implies & k \cdot (N + 1 - p - q) + 1 = 0 \pmod{e} \end{aligned}$$

Here the unknowns are k and $(-p - q)$. We can write that problem as a polynomial with root x_0 and y_0 :

$$f(x, y) = x \cdot (A + y) \pmod{e}$$

with $A = N + 1$ and $y = -p - q$.

Now we use **Coppersmith's heuristic for multivariate polynomials**. Coupled with **Howgrave-Graham's Theorem** for bivariate polynomials :

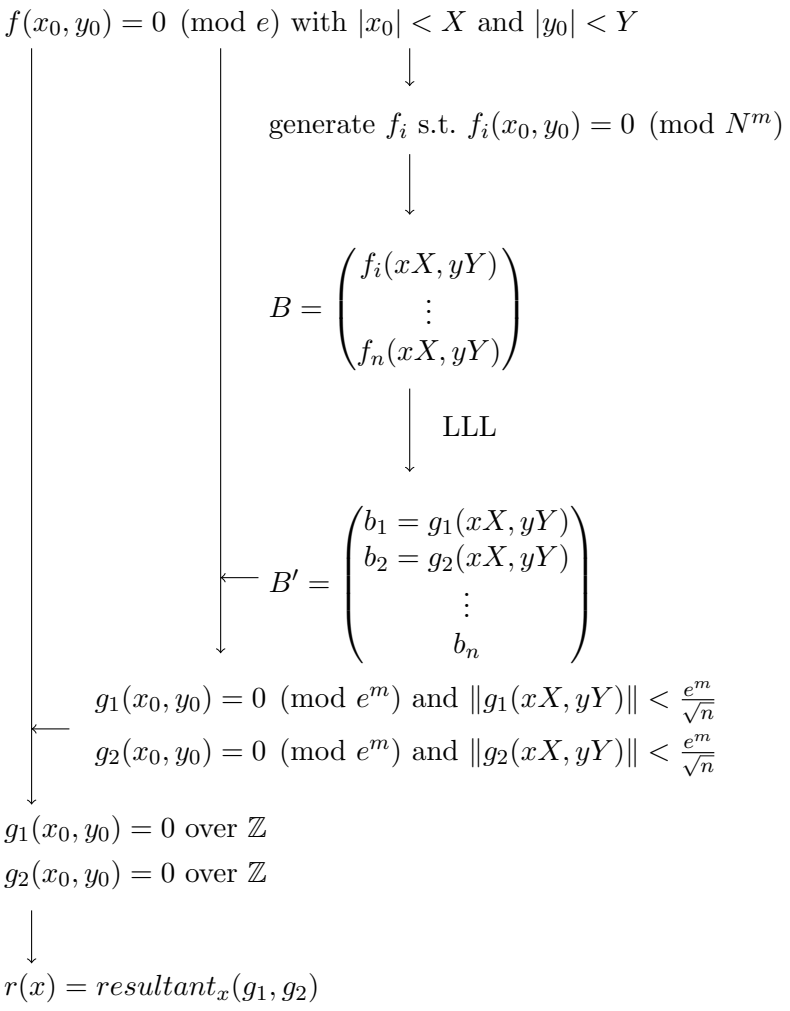
Theorem 5. *Let $g(x)$ be an bivariate polynomial with at most n monomials. Further, let m be a positive integer. Suppose that*

$$g(x_0, y_0) = 0 \pmod{e^m} \quad \textit{where} \quad |x_0| \leq X \textit{ and } |y_0| \leq Y \tag{1}$$

$$\|g(xX, yY)\| < \frac{e^m}{\sqrt{n}} \tag{2}$$

Then $g(x_0, y_0) = 0$ holds over the integers.

But the problem here is that **one polynomial is not enough** to get the roots of a bivariate equation. What we need are **two polynomials**, then we could use the **resultant** or a Gröbner basis to find the roots.
Coppersmith proposed to take **the two shortest vectors**, of the LLL-reduced basis, as polynomials. Let's take a look at what it should look from a distance :



And once we find the root x_0 of r we can re-inject it in g_1 to find y_0.
This doesn't always yield a solution. For example, if the two polynomials g_1 and g_2 are not independent, the resultant will be zero.

Boneh and Durfee proposed a construction of the f_i polynomials as follow :

$$\text{for } k = 0, \ldots, m :$$
$$g_{i,k}(x) = x^i \cdot f^k(x, y) \cdot e^{m-k} \text{ for } i = 0, \ldots, m - k$$
$$h_{j,k}(x) = y^j \cdot f^k(x, y) \cdot e^{m-k} \text{ for } j = 0, \ldots, t$$

They called the $g_{i,k}$ **x-shifts** and the $h_{j,k}$ **y-shifts**.

By using these polynomials to build the lattice, carefully balancing the variables so that the determinant of the triangular basis doesn't exceed e^{mn},

Boneh and Durfee showed that LLL successfuly yielded useful results if $d < N^{0.284}$.

To achieve their **improved results** of $d < N^{0.292}$, they showed that using a sublattice by **ignoring some of the y-shifts**, the bounds on the shortest vectors found by LLL were improved.
This is because of the "**helpful vectors**" notion of Howgrave-Graham. A vector is helpful if his contribution to the determinant (its monomial that appears in the diagonal of the lattice basis) is less than e^m. Boneh and Durfee's method is to discard all y-shifts when their highest monomial exceeds e^m.

	1	x	xy	x^2	x^2y	x^2y^2	y	xy^2	x^2y^3
e^2	e^2								
xe^2		e^2X							
fe	e	eAX	eXY						
x^2e^2				e^2X^2					
xfe		eX		eAX^2	eX^2Y				
f^2	1	$2AX$	$2XY$	A^2X^2	$2AX^2Y$	X^2Y^2			
ye^2						e^2Y			
yfe			$eAXY$				eY	eXY^2	
yf^2			$2AXY$		A^2X^2Y	$2AX^2Y^2$	Y	$2XY^2$	X^2Y^3

Boneh-Durfee basis matrix for $m = 2$, $t = 1$

	1	x	xy	x^2	x^2y	x^2y^2	y	xy^2	x^2y^3
e^2	e^2								
xe^2		e^2X							
fe	e	eAX	eXY						
x^2e^2				e^2X^2					
xfe		eX		eAX^2	eX^2Y				
f^2	1	$2AX$	$2XY$	A^2X^2	$2AX^2Y$	X^2Y^2			
yf^2			$2AXY$		A^2X^2Y	$2AX^2Y^2$	Y	$2XY^2$	X^2Y^3

After removing the damaging y-shifts' coefficient vectors

Unfortunately by doing this we **lose the triangular structure** of the basis and evaluating the determinant of the new rectangular basis is tricky.
Boneh and Durfee developed the notion of **Geometrically progressive matrices** to handle these non-triangular lattice basis. Later **Blömer and May** took a different approach by noticing that some of the columns could be removed without affecting the determinant too much, that allowed the lattice basis to return to a triangular structure. Both those methods are quite difficult to handle and it's more recently that **Herrmann and May** found a clever and better way :

The **Unravelled Linearization** technique consists in **modifying** our initial polynomial f. Done cleverly this will modify the f_i so that after removing the y-shifts, our sublattice basis will **directly be triangular**.
Herrmann and May propose to do the following **substitution** on f :

$$f(x,y) = \underbrace{1+xy}_{u} + Ax \pmod{e}$$

This leaves us with the **linear polynomial** $\bar{f}(u,x) = u + Ax$ (note the lexicographic order, u before x) and a relation $xy = u - 1$.
The x-shifts are still constructed as usual :

$$\bar{g}_{i,k}(u,x) = x^i \cdot \bar{f}^k \cdot e^{m-k} \text{ for } k = 0,\dots,m \text{ and } i = 0,\dots,m-k$$

The y-shifts are constructed the same way, but we need to apply our relation again afterward to completely perform our unravelled linearization :

$$\bar{h}_{j,k}(u,x,y) = y^j \cdot \bar{f}^k \cdot e^{m-k} \text{ for } j = 1,\dots,t \text{ and } k = \left\lfloor \frac{m}{t} \right\rfloor \cdot j,\dots,m$$

They are **selected** with the notion of "**increasing pattern**" in mind, so that using the previous relation $xy = u - 1$ we end up with a **triangular** lattice basis :

$$\begin{array}{c} \\ e^2 \\ xe^2 \\ \bar{f}e \\ x^2e^2 \\ x\bar{f}e \\ \bar{f}^2 \\ y\bar{f}^2 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & x & u & x^2 & ux & u^2 & u^2y \end{array} \\ \begin{pmatrix} e^2 & & & & & & \\ & e^2X & & & & & \\ & eAX & eU & & & & \\ & & & e^2X^2 & & & \\ & & & eAX^2 & eUX & & \\ & & & A^2X^2 & 2AUX & U^2 & \\ & -A^2X & -2AU & & A^2UX & 2AU^2 & U^2Y \end{pmatrix} \end{array}$$

The same matrix as above after unravelled linearization

Now that we have built a lattice basis, we have to know how we need to **bound our two shortest vectors** so that Howgrave-Graham second point is respected. From LLL's property :

$$\|v_1\| \leq \|v_2\| \leq 2^{\frac{n}{4}} \cdot det(L)^{\frac{1}{n-1}}$$

We need to bound v_1 and v_2 so that they respect Howgrave-Graham's theorem second point. This is the bound we end up with on the determinant :

$$det(L) < \frac{e^{m(n-1)}}{(n2^n)^{\frac{n-1}{2}}}$$

That can be reduced by removing the "error terms" :

$$det(L) < e^{mn}$$

6.2 How was the bound on d calculated ?

Let's go back to our first equation :

$$e \cdot d = k \cdot \varphi N + 1 = k \cdot (N + 1 - p - q)$$

Let's recap what we know :
– p and q should be half the size of N :

$$\frac{log(p)}{log(N)} \approx \frac{1}{2}$$
$$\implies p, q \approx N^{\frac{1}{2}}$$

– $e \approx N$
– $d < N^{\delta}$

The first two are what we usually use in RSA. Our last one is the bound we are trying to define. Of course we want a δ as small as possible. Now that we have defined all these, let's go back to our previous equation and let's bound our unknowns :

$$k \cdot (\underbrace{N+1}_{A} + \underbrace{-p-q}_{s}) \pmod{e} \quad \text{with} \begin{cases} |s| \approx 2N^{\frac{1}{2}} \\ |k| < \frac{N^{\delta}-1}{N-1+2N^{\frac{1}{2}}} \end{cases}$$

s should be pretty close to our prediction. But k ? It could be way lower, this incertitude tells us we have the possibility to play with its bound.
Let's note that $s >> k$ and that reducing the size of our unknown s is a good idea. Notice that both s and A are even, that allows us to reformulate our problem with a smaller s :

$$f(x, y) = x \cdot (A + y) \pmod{e} \quad \text{with} \begin{cases} x = 2k \\ y = (-p-q)/2 \\ A = (N+1)/2 \end{cases}$$

We can now reformulate our bounds and remove the negligible terms :

$$\begin{cases} |y| \approx N^{\frac{1}{2}} \\ |x| < 2N^{\delta} \end{cases}$$

We do not need to bound them accurately as the LLL property on the shortest vector bound is a dramatic one.

Now for our algorithm we need to fix a m and a t such that $dim(L) = n$. For Herrmann and May's selection of the y-shifts we need $m >= t$ so we will write $t = \tau m$ with $\tau < 1$.
Remember earlier we said that we need to bound our determinant to obtain **Howgrave-Graham's Theorem second point**, and **LLL's property** gave us this **bound** :

$$det(L) < e^{mn}$$

$det(L)$ being a function of $e \approx N$, δ, m and t. The bound being a function of e, m and t.
from Herrmann and May's proof we end up with this formula as **determinant** (after removing negligible terms that were calculated for $m \to \infty$) :

$$det(L) = X^{\frac{1}{6}m^3} Y^{\frac{\tau^2}{6}m^3} U^{(\frac{1}{6}+\frac{\tau}{3})m^3} e^{(\frac{1}{3}+\frac{\tau}{2})m^2}$$

with this dimension of the lattice :

$$dim(L) = n = \left(\frac{1}{2} + \frac{\tau}{2}\right) m^2$$

We then **replace the upper bounds** by their values : $X = N^{\delta}$, $Y = N^{\frac{1}{2}}$, $U = X * Y$ and we see for what m, τ and δ the algorithm still runs in polynomial time (the complexity is dominated by LLL's).
We end up with an optimized value of $\tau = (1 - 2\delta)$ and the **Boneh-Durfee bound** :

$$\delta \leq \frac{1}{2}(2 - \sqrt{2}) \approx 0.292$$

6.3 Experiments

My experiments were far from Boneh and Durfee's findings. When they took hours to solve for small m and t, I took seconds. This is, I guess, the gap between 1999 and 2015 : I used Sage 6.4 in a Virtualbox with 512Mo of RAM and 1 core from an Intel i7 @ 2.30GHz. I found out that in practice, the bound of X was often way higher than the root x_0, decreasing X until the algorithm worked was a good way to find the roots. Also, the LLL-reduced basis first two vectors were not always independent but I could often finds two independent vectors in the reduced basis.

δ	size of N (bits)	size of d (bits)	m	t	running time
0.26	1024	256	3	1	0.8s
0.26	2048	532	3	1	1.9s
0.27	2048	553	6	3	3m 3s
0.28	1024	553	6	3	3m 3s

7 Summary

Références

[1] Chris Peikert *Lattices in Cryptography, Georgia Tech, Fall 2013 : Lecture 2, 3*

[2] Don Coppersmith *Finding Small Solutions to Small Degree Polynomials*

[3] Nicholas Howgrave-Graham *Finding Small Roots of Univariate Modular Equations Revisited*

[4] Alexander May *Using LLL-Reduction for Solving RSA and Factorization Problems*

[5] Boneh and Durfee *Cryptanalysis of RSA with Private Key d Less Than* $N^{0.292}$

[6] Herrmann and May *Maximizing Small Root Bounds by Linearization and Applications to Small Secret Exponent RSA*