Cryptanalysis of a public-key cryptosystem
based on approximations by rational numbers *

Jacques Stern
Équipe de Logique
Université de Paris 7
et
Département de mathématiques et informatique
École Normale Supérieure

Philippe Toffin
Département de mathématiques
Université de Caen

Abstract

At the Eurocrypt meeting, a public-key cryptosystem based on rational numbers has been proposed [2]. We show that this system is not secure. Our attack uses the LLL algorithm. Numerical computations confirm that it is successful.

1 The proposed cryptosystem

We briefly review the article of H.Isselhorst [2], in which the system was described. The secret key consists of a large prime number p, (a size of 250 decimal digits is suggested), together with a (small) integer k and a (k,k)-matrix A, with an inverse A^{-1} mod p.

The public part of the system essentially consists of a matrix

$$C = (c_{i,j})_{1 \leq i \leq k \;\&\; 1 \leq j \leq k}$$

where $c_{i,j}$ is computed from A and a fixed public integer t, $1 \leq t < p$, by truncating the decimal expansion of $ta_{i,j}/p$ after n digits, which we write

$$c_{i,j} = \text{Float}(ta_{i,j}/p, n)$$

Also included in the public key data are integers z and m satisfying inequalities which will be given later on

The plaintext is a vector X, with k coordinates, all of them being positive integers bounded by m. The encryption is as follows:

*Research supported by the PRC mathématiques et informatique
I.B. Damgard (Ed.): Advances in Cryptology - EUROCRYPT '90, LNCS 473, pp. 313-317, 1991.

- Compute $U = C.X$
- Set $V = U \bmod t$, where the mod function is applied coordinatewise
- Output $Y = \mathrm{Float}(V, z)$, where the Float function is applied coordinatewise

We now explain why and how the ciphertext can be decoded. If u_i (resp. v_i) is the ith coordinate of U (resp. V), we can write:

$$\frac{pv_i}{t} = \frac{pu_i}{t} \bmod p$$

and using the definition of Y,

$$\frac{py_i}{t} = \frac{pv_i}{t} - \frac{pe_i}{t} \text{ with } 0 \leq e_i < 10^{-z}$$

thus, for some integers α_i, we get

$$\frac{py_i}{t} = \alpha_i p + \frac{pv_i}{t} - \frac{pe_i}{t}$$

which gives

$$\frac{py_i}{t} = \alpha_i p + \frac{p}{t}\sum_j c_{i,j}x_j - \frac{pe_i}{t}$$

$$= \alpha_i p + \frac{p}{t}\sum_j \Big(\frac{ta_{i,j}}{p} - r_{i,j}\Big)x_j - \frac{pe_i}{t}$$

$$= \alpha_i p + \sum_j a_{i,j}x_j - \frac{p}{t}\sum_j r_{i,j}x_j - \frac{pe_i}{t}$$

Noting that

$$0 \leq r_{i,j} < 10^{-n}$$

we get that the sum of the last two terms is bounded by

$$\frac{p}{t}10^{-n}km + \frac{p}{t}10^{-z}$$

Now, if both terms are bounded by $1/4$, then the real value of

$$\sum_j a_{i,j}x_j$$

can be easily recovered from p, by rounding py_i/t and reducing mod p. From these values, the original message is obtained via A^{-1}

The inequalities that are needed to carry through the above argument are easy consequences of thoses which are proposed in the paper [2], namely

$$10^{70} \leq m \leq p/10$$

$$4kpm/t \leq 10^n < p^2 10^{-50}/t$$

$$10^{z-1} \leq 4p/t < 10^z$$

Presumably, the other inequalities have been added to ensure security.

2 A cryptanalytic attack

Our attack uses the LLL algorithm [3], very much in the same way as known attacks against the knapsack-based cryptosystem (see [1]). It can be described as follows

• Pick four distincts values c_1, c_2, c_3, c_4 among the k^2 possible $c_{i,j}$'s, included the largest one c_1. Set

$$\gamma_i = 10^n c_i \quad 1 \leq i \leq 4$$

Note that the γ_i's are integers.

• Apply the LLL algorithm to the 4-dimensional lattice generated by the columns of the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\gamma_2 & \gamma_1 & 0 & 0 \\ -\gamma_3 & 0 & \gamma_1 & 0 \\ -\gamma_4 & 0 & 0 & \gamma_1 \end{pmatrix}$$

• Output the first coordinate a_1 of the first vector of the reduced basis of L, obtained through LLL.

We claim that a_1 is precisely the original value $a_{i,j}$, corresponding to the largest of the $c_{i,j}$, which was choosen as c_1. We will give a heuristic justification of this fact. The argument can actually be put on a firmer theoretical basis by a precise probabilistic analysis. Anyhow, as will be seen in section 3, the success of the attack is confirmed by numerical experiments.

First observe that, for $i = 1,..4$, if we denote by $a_1,..a_4$ the values of $a_{i,j}$ corresponding to $c_1,..c_4$, we have

$$0 \leq \frac{ta_i}{p} - c_i < 10^{-n}$$

which gives , by linear combination

$$|a_1 c_i - a_i c_1| \leq 10^{-n} p, \quad i = 2, 3, 4$$

multiplying by 10^n, we get

$$|a_1 \gamma_i - a_i \gamma_1| \leq p, \quad i = 2, 3, 4$$

together with the inequality $1 \leq a_1 < p$, this shows that the integers a_1, a_2, a_3, a_4 provide a linear combination V of the columns of the matrix of L, whose coordinates are bounded by p. Now, the determinant of L is γ_1^3. Because c_1 is the largest of the $c_{i,j}$'s, it is presumably close to t; thus γ_1 is close to $t10^n$ so that the expected size of the coordinates of a short vector is about $3(n + \log_{10} t)/4$ digits. Letting $m = p^\alpha$, and using the fact that

$$4kpm/t \leq 10^n$$

we see that the expected value of the coordinates of a short vector of L should be bounded from below by $p^{3(1+\alpha)/4}$. If α is significantly greater than $1/3$, $p^{3(1+\alpha)/4}$ is definitely greater than p, so that the LLL algorithm will actually disclose the very short vector V, defined above, whose first coordinate is precisely a_1.

If we consider the size suggested in [2], namely 250 digits, we see that our attack is presumably successful when the size of the coded messages m is 85 digits or more. Of corse, for a smaller choice of m, it is possible to apply an analogous method, provided one chooses more than 4 values c_i and one apply the LLL algorithm in a larger dimension. For example, the 6D-version of the attack works as soon as α is significantly greater than 1/5.

Once a_1 has been correctly recovered, p can be computed by rounding ta_1/c_1; this because of the inequality

$$\left|\frac{ta_1}{c_1} - p\right| \leq \frac{p}{c_1 10^n} = \frac{p}{\gamma_1}$$

Similarly, the correct value of each $a_{i,j}$ is obtained by rounding $a_1 c_{i,j}/c_1$. This is because of the following inequality

$$\left|a_{i,j} - \frac{a_1}{c_1} c_{i,j}\right| \leq \frac{p}{c_1 10^n} = \frac{p}{\gamma_1}$$

3 Numerical experiments

For numerical experiments, we used the Symbolic Computation System Maple. In all our experiments, we restricted ourselves to the case $k = 2$, which involves a (4,4)-matrix Mat to be reduced by LLL.

3.1 Main part of our program

In order to test our cryptanalytic attack, we first choose the 3 following parameters: t, n , whose role is explained in the previous sections and nb_of_digits which is the number of digits of the prime number p. We then choose randomly 4 nb_of_digits-long integers a_i, a_1 beeing the largest, and p a prime number greater than these 4 numbers. The c_i are the values of $Float(ta_i/p, n)$ and are the public key. The γ_i and the matrix Mat are then built as in section 2 and we obtain by LLL-reduction a new matrix new_Mat. We may assume that $new\_Mat_{1,1}$ is positive. Our algorithm fails if $new\_Mat_{1,1}$ is different of a_1, and if not, we let $new\_a_1 = new\_Mat_{1,1}$. We then get a value

$$new\_p = \text{closest\_integer}(t \ * \ 10^n \ new\_a_1/\gamma_1)$$

Again the attack fails if $new\_p \neq p$, and if not, we let

$$new\_a_i = \text{closest\_integer}(new\_a_1 \ * \ \gamma_i/\gamma_1)$$

We reach complete success if for each i, we have $new\_a_i = a_i$.

3.2 Results

We made 4 different trials under different values of the parameters.

• $t = 1$, $nb\_of\_digits = 20$, $n = 30$: 10 different runs reached complete success.

• The smallest bound for m suggested by Isselhorst beeing 70, when $t = 1$, we took $nb\_of\_digits = 121$, and $n = 192$: 2 different runs reached complete success.

• $nb\_of\_digits = 121$, and t is a randomly choosen 50-digits integer. As is clear from section 2, this allows a lower value for n. We set $n = 141$, which corresponds to messages m of length 70. 2 different runs reached complete success.

• The runs with the largest figures:
$t = 1$, $nb\_of\_digits = 250$, $n = 336$, which allows messages m of length 85. 4 different runs reached complete success.

3.3 Remarks

All our trials gave values which can be saved and can be used again. We also checked that for $t = 1$, $nb\_of\_digits = 20$, we get a failure as soon as $n \leq 27$. This is in accordance with the analysis os section 2.

3.4 Final conclusion

All these complete success justify the basic claim of our theoretical analysis above: The cryptosystem proposed by Isselhorst is not secure.

References

[1] E.F.Brickell, The cryptanalysis of knapsack cryptosystems. *Proceedings of the third SIAM discrete mathematics conference.*

[2] H.Isselhorst, The use of fractions in public-key cryptosystems. *Proceedings Eurocrypt'89.*

[3] A.K.Lenstra, H.W.Lenstra, L.Lovász, Factoring polynomials with rational coefficients. *Math. Annalen* 261 (1982) 515-534.