On The Complexity Of Finding Low-Level Solutions

Björn Grohmann

nn@mhorg.de

1 Introduction

This is the second part of the authors' article *An Applicable Public-Key-Cryptosystem Based On NP-Complete Problems* (cf. [2]), where it was shown that the security of the proposed PKC mainly relys on the expected hardness of finding a special kind of solution $(\mathbf{x}, \mathbf{y}, \lambda) \in \mathbb{F}_p^m \times \mathbb{F}_p^n \times \mathbb{F}_p$ of the equation

$$A\mathbf{x} + \lambda\mathbf{y} = \mathbf{b}, \tag{1}$$

with, for a prime $p > 2$, $\mathbb{F}_p$ being a finite field with p elements, $A \in \mathbb{F}_p^{n\times m}$ a matrix and $\mathbf{b} \in \mathbb{F}_p^n$ a vector.

More specifically, it was shown in [2] that a necessary condition for the existence of an efficient decoding algortihm for the proposed PKC is that a solution $(\mathbf{x}, \mathbf{y}, \lambda)$ of equation (1) has **level** t, for a "small" integer t, and it has been proven that for a large part of the class of these "low-level solutions", their computation is in general a NP-complete task.

The aim of this article is to prove the following two theorems:

Theorem 1 *Let $t > 0$ be an integer constant. Given a prime $p > 2$, positive integers n, m and a solution $(\mathbf{x}, \mathbf{y}, \lambda) \in \mathbb{F}_p^m \times \mathbb{F}_p^n \times \mathbb{F}_p$ having level t. Then there exists an integer c, only depending on t, and a representation of the vectors $\mathbf{x}$ and $\mathbf{y}$ of the form $\mathbf{x} = \sum_{i=1}^{l} \alpha_i \mathbf{x_i}$ and $\mathbf{y} = \sum_{i=1}^{l} \beta_i \mathbf{y_i}$, with $l = \lfloor \log^c(nm) \rfloor$ and $\alpha_i, \beta_i \in \mathbb{F}_p$, $\mathbf{x_i} \in \{0,1\}^m$, $\mathbf{y_i} \in \{0,1\}^n$, for $i = 1, \dots, l$.*

Theorem 2 *Let $c \geqslant 0$ be an integer constant. Given a prime $p > 2$, positive integers n, m, a matrix $A \in \mathbb{F}_p^{n\times m}$ and a vector $\mathbf{b} \in \mathbb{F}_p^n$. Deciding, whether there exists an element $\lambda \in \mathbb{F}_p$ and vectors $\mathbf{x} = \sum_{i=1}^{l} \alpha_i \mathbf{x}_i$ and $\mathbf{y} = \sum_{i=1}^{l} \beta_i \mathbf{y}_i$, with $l = \lfloor \log^c(nm) \rfloor$, $\alpha_i, \beta_i \in \mathbb{F}_p$, $\mathbf{x}_i \in \{0,1\}^m$, $\mathbf{y}_i \in \{0,1\}^n$, for $i = 1, \dots, l$, such that $A\mathbf{x} + \lambda\mathbf{y} = \mathbf{b}$, is NP-complete.*

2 The Complexity of Low-Level Solutions

We start by fixing some notation. Let $\mathbb{Z}$ be the set of integers. For a prime $p > 2$, the finite field with p elements will be denoted by $\mathbb{F}_p$ and its subgroup of non-zero elements by $\mathbb{F}_p^\times$. We will use a representation of elements of $\mathbb{F}_p$ of the form $\mathbb{F}_p = \{-(p-1)/2, \ldots, (p-1)/2\}$ and we will frequently view integers as elements of $\mathbb{F}_p$ and vice versa, if the context allows this. All vectors $\mathbf{x} \in \mathbb{F}_p^n$ will be viewed as column vectors, the transpose of a vector $\mathbf{x}$ will be denoted by $\mathbf{x}^\mathsf{T}$. For two vectors $\mathbf{x} = (x_i)_i^\mathsf{T}$ and $\mathbf{y} = (y_i)_i^\mathsf{T}$ we denote their (inner) product by $\mathbf{x}^\mathsf{T}\mathbf{y} = \sum_i x_i y_i$. For two integers s and t, with $s \leqslant t$, we will write $\langle s, t\rangle^n$ to denote the set of vectors $\mathbf{x}^\mathsf{T} = (x_1, \ldots, x_n) \in \mathbb{Z}^n$ with $s \leqslant x_i \leqslant t$, for $i = 1, \ldots, n$, so, by abuse of notation, $\mathbb{F}_p^n = \langle -(p-1)/2, (p-1)/2\rangle^n$. Finally, the number of elements of a finite set $\mathcal{S}$ will be denoted by $|\mathcal{S}|$.

Let us first recall two definitions from [2]: for a vector $\mathbf{x} \in \mathbb{F}_p^m$ of finite dimension $m > 0$ we define a counting function κ via

$$\kappa(\mathbf{x}) = \left|\left\{\mathbf{x}^\mathsf{T}\mathbf{z} \mid \mathbf{z} \in \{0,1\}^m\right\}\right|, \tag{2}$$

that is the number of different values of sums of all possible subsets of components of $\mathbf{x}$. It can be shown (cf. [2]) that (for finite dimensions) the computation of the exact value of κ is in general NP-hard and coNP-hard, but nevertheless, at least for vectors of a special kind, its value can be reasonably bounded from above.

Next, we define the level of a solution. For that, let t, m and n denote positive integers. We say that a solution $(\mathbf{x}, \mathbf{y}, \lambda) \in \mathbb{F}_p^m \times \mathbb{F}_p^n \times \mathbb{F}_p$ has **level** t, if

$$(nm)^{t-1} < \max\left(\kappa(\mathbf{x}), \kappa(\lambda\mathbf{y}), \kappa(\mathbf{x})\kappa(\lambda\mathbf{y})\right) \leqslant (nm)^t. \tag{3}$$

Proof of Theorem 1. First we note that it is enough to prove the existence of the claimed representation for a single vector $\mathbf{x} \in \mathbb{F}_p^n$. So let t be a positive integer and let us assume that $n^{t-1} < \kappa(\mathbf{x}) \leqslant n^t$. Then, by definition of κ, there exists an integer $k \geqslant 2^n/n^t$ and a submatrix $M' \in \mathbb{F}_p^{k\times n}$ of the matrix $M \in \mathbb{F}_p^{2^n \times n}$ of all bit-strings of length n, and an element $\gamma \in \mathbb{F}_p$ such that $M'\mathbf{x} = \gamma\mathbf{1}$, where $\mathbf{1}$ denotes the all-one vector. If $\gamma \neq 0$, we build a new matrix $M'' \in \mathbb{F}_p^{k\times n}$ by replacing each row of M' with the vector obtained by substracting this row from the first row of M', which yields $M''\mathbf{x} = \mathbf{0}$. Now, for n large enough, the rank of M'' is greater than $\log(2^n/n^{t+1}) = n - (t+1)\log(n)$ which means that the dimension of its kernel is at most $(t+1)\log(n)$ and therefore the vector $\mathbf{x}$ has a representation of the form $\mathbf{x} = \sum_{i=1}^{\lfloor (t+1)\log(n)\rfloor} \alpha_i \mathbf{x_i}$, with $\alpha_i \in \mathbb{F}_p$ and $\mathbf{x_i} \in \langle -n^s, n^s\rangle^n$, for $i = 1, \ldots, \lfloor (t+1)\log(n)\rfloor$ and some constant s, so, by picking a basis $\beta_0 = 1, \beta_1 = 2, \ldots, \beta_{\lfloor s\log(n)\rfloor} = 2^{\lfloor s\log(n)\rfloor}$ for the set $\{1, \ldots, n^s\}$, Theorem 1 follows. $\square$

Proof of Theorem 2. The proof of this theorem needs a little bit of preparation. Let l be a positive integer and denote by M a matrix of dimension $(l+1) \times l$ with entries from the set $\{0, 1\}$. Further, let $\mathbf{d} = (d_1, \ldots, d_{l+1})^T \in \mathbb{F}_p^{l+1}$ be a vector with $d_1 = 1$ and, for $k = 1, \ldots, l-1$,

$$d_{k+1} = l^l \sum_{j=1}^{k} d_j, \tag{4}$$

and finally $d_{l+1} = \sum_{j=1}^{l} d_j$. We now claim that, for p large enough, a vector $\mathbf{x} = (x_1, \ldots, x_l)^T \in \mathbb{F}_p^l$ is a solution of the equation $M\mathbf{x} = \mathbf{d}$ if and only if there exists a permutation π on the set $\{1, \ldots, l\}$ such that $x_i = d_{\pi(i)}$, for $i = 1, \ldots, l$. To see this, denote by M' the $l \times l$ submatrix of M where the last row has been deleted. Equivalently, we denote by $\mathbf{d}' = (d_1, \ldots, d_l)^T$. Now, a solution $\mathbf{x}$ of the equation $M'\mathbf{x} = \mathbf{d}'$ exists if and only if the rank of M' = the rank of $(M'|\mathbf{d}')$ and therefore it follows that $\det(M') \neq 0$, by definition of the d_i. Please note further that for $l > 1$, the determinant of M' (viewed over $\mathbb{Z}$) is clearly less than l^{l-1} and that (again viewed over $\mathbb{Z}$) for every sum $\sum_{j=1}^{l} a_j d_j = 0$, with $|a_j| < l^{l-1}$, we have $a_j = 0$ for all j. So, we can conclude that the last row of M is the all-one vector and that M' has to be a permutation matrix.

For the next step, let F be a Boolean function. It is well known (cf. [4]) that there exists a function F' that is satisfiable if and only if F is satisfiable, and which can be written in the following form:

$$F' = x_0 \wedge (a_1 \leftrightarrow (b_1 \circ c_1)) \wedge \cdots \wedge (a_t \leftrightarrow (b_t \circ c_t)), \tag{5}$$

for a positive integer t, where x_0 is a variable, a_i, b_i, c_i are literals and $\circ \in \{\wedge, \vee\}$. Clearly, the two types of terms can be written as

$$(a \leftrightarrow (b \vee c)) = (a \vee \neg b) \wedge (\neg a \vee b \vee c) \wedge (a \vee \neg c) \tag{6}$$

$$(a \leftrightarrow (b \wedge c)) = (\neg a \vee b) \wedge (a \vee \neg b \vee \neg c) \wedge (\neg a \vee c) \tag{7}$$

and the reason why we recall this rather elementary fact is to point out that if F' is satisfiable, then at most two of the literals in each clause can have a TRUE-assignment. We further assume that each variable of F' appears at most once in each clause.

Next, we define a matrix A' of dimension $(3t + 1) \times t'$, where t' is the number of different variables of F', such that if the variable "x_j" appears in clause i, we put a "1" at the position (i, j), except when the clause is of the form $(x_j \vee \neg x_s \vee \neg x_t)$, where we put a "2" at position (i, j) of A'. Else, if "$\neg x_j$" is in clause i, we put a "-1" at position (i, j) and if the variable "x_j" is not part of clause i, we put a "0" at position (i, j) of A'.

The final matrix A is now of the form

$$A = \left(\begin{array}{c|c} A' & 0 \\ \hline 0 & I_{l+1} \\ \hline 0 & 0 \end{array} \right) \in \mathbb{F}_p^{(3t+1+l+1+l+1)\times(t'+l+1)}, \tag{8}$$

where I_{l+1} denotes the identity matrix of dimension $l+1$ for a positive integer l. Now, let $\mathbf{d} = (d_1, \ldots, d_{l+1})^T \in \mathbb{F}_p^{l+1}$ be the vector from above. We will define our vector $\mathbf{b} \in \mathbb{F}_p^{3t+1+l+1+l+1}$ as follows. The first $3t+1$ components depend on the shape of the clauses of F' in a sense that, if the i-th clause has one of the forms (x_j), $(x_j \vee x_s)$ or $(x_j \vee x_s \vee x_t)$, for variables x_j, x_s, x_t of F', then we define the i-th component of $\mathbf{b}$ to be "$2d_{l+1}$". If the i-th clause of F' has one of the forms $(x_j \vee \neg x_s)$, $(\neg x_j \vee x_s \vee x_t)$ or $(x_j \vee \neg x_s \vee \neg x_t)$, then the i-th component of $\mathbf{b}$ is "d_{l+1}". If the i-th clause of F' has the form $(\neg x_j \vee \neg x_s)$, then the i-th component of $\mathbf{b}$ is "0", and finally, if the i-th clause of F' has the form $(\neg x_j \vee \neg x_s \vee \neg x_t)$, then the i-th component of $\mathbf{b}$ is defined to be "$-d_{l+1}$". The last $2(l+1)$ components of $\mathbf{b}$ will be two copies of the vector $\mathbf{d}$.

It is now an easy exercise to verify that the function F' (resp. F) is satisfiable, if and only if a solution $(\mathbf{x}, \mathbf{y}, \lambda)$ of the equation $A\mathbf{x} + \lambda\mathbf{y} = \mathbf{b}$ and of the required form exists. If F' is satisfiable and the i-th variable has an assignment TRUE (resp. FALSE), then putting the i-th component of $\mathbf{x}$ to be "d_{l+1}" (resp. "0") leads to a valid solution. On the other hand, if $(\mathbf{x}, \mathbf{y}, \lambda)$ is a solution of the required form, then if the i-th component of $\mathbf{x}$ is in the set $\{1, \ldots, d_{l+1}\}$, defining the i-th variable of F' to be TRUE (else FALSE) shows that F' is satisfiable. $\square$

References

[1] McEliece, J.R.: A public-key cryptosystem based on algebraic coding theory. In: Deep Space Network Progress Report 42-44, Jet Propulsion Lab, California Institute of Technology, pp. 114-116 (1978)

[2] Grohmann, B.: An Applicable Public-Key-Cryptosystem Based On NP-Complete Problems. Report 2014/117, eprint.iacr.org (2014)

[3] Neukirch, J.: Algebraische Zahlentheorie. Springer, Berlin, Heidelberg (1992)

[4] Schoening, U.: Theoretische Informatik kurz gefasst. BI-Wissenschaftsverlag, Mannheim (1992)

[5] Shor, P.W.: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. arXiv:quant-ph/9508027 (1995)