On the Modular Inversion Hidden Number Problem

San Ling

Div. of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, 637371

Igor E. Shparlinski

Dept. of Computing, Macquarie University, Sydney, NSW 2109, Australia

Ron Steinfeld

Dept. of Computing, Macquarie University, Sydney, NSW 2109, Australia

Huaxiong Wang

Div. of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371

Abstract

We give a rigorous deterministic polynomial time algorithm for the modular inversion hidden number problem introduced by D. Boneh, S. Halevi and N. A. Howgrave-Graham in 2001. For our algorithm we need to be given about 2/3 of the bits of the output, which matches one of the heuristic algorithms of D. Boneh, S. Halevi and N. A. Howgrave-Graham and answers one of their open questions. However their more efficient algorithm that requires only 1/3 of the bits of the output still remains heuristic.

Key words: Hidden number problem, inversion, pseudorandom generators

Email addresses: lingsan@ntu.edu.sg (San Ling), igor.shparlinski@mq.edu.au (Igor E. Shparlinski), ron.steinfeld@mq.edu.au (Ron Steinfeld), hxwang@ntu.edu.sg (Huaxiong Wang).

Preprint submitted to Elsevier *8 March 2011*

1. Introduction

1.1. Motivation

Since Boneh and Venkatesan [6,7] introduced the *hidden number problem* in 1996, it has been generalised in a number of directions and has found a wide spectrum of applications in cryptography and beyond, see [16] for a survey of relevent results and also [1] for some recent developments and a new approach. Here we consider a modification of the original problem which has been introduced by Boneh, Halevi and Howgrave-Graham [5].

More precisely, for a prime p, denote by $\mathbb{F}_p$ the field of p elements and always assume that it is represented by the set $\{0, 1, \ldots, p-1\}$. Accordingly, sometimes, where obvious, we treat elements of $\mathbb{F}_p$ as integer numbers in the above range.

For integers x and $m \geq 1$ we denote by $\lfloor x \rfloor_m$ the remainder of x on division by m. Given an integer m and $\ell > 0$, we denote by $\mathrm{MSB}_{\ell,m}(x)$ any integer u such that

$$|\lfloor x \rfloor_m - u| \leq m/2^{\ell+1}.$$

Roughly speaking, $\mathrm{MSB}_{\ell,m}(x)$ gives ℓ most significant bits of the remainder of x modulo m. However, this definition is more flexible and suits better our purposes. In particular we remark that ℓ in the above inequality need not be an integer.

Following Boneh, Halevi and Howgrave-Graham [5] we consider the following *Modular Inversion Hidden Number Problem*, **ModInv-HNP**:

Recover a number $\alpha \in \mathbb{F}_p$ such that for N elements $t_1, \ldots, t_N \in \mathbb{F}_p \setminus \{-\alpha\}$, chosen independently and uniformly at random, we are given N pairs

$$\left(t_i, \mathrm{MSB}_{\ell,p}\left(\frac{1}{\alpha + t_i}\right)\right), \qquad i = 1, \ldots, N,$$

for some $\ell > 0$.

Besides being of independent interest and giving an interesting example of yet another natural problem of this type, see [16] for a survey, it has also been mentioned in [5] as a building block for constructing efficient pseudorandom number generators and message authentication codes. Motivated by these applications, here we modify and rigorously analyze the algorithm outlined in [5, Section 3.1], using some ideas from [2–4]. We note that our algorithm works only if for some fixed $\varepsilon > 0$ we have $\ell > (2/3 + \varepsilon)k$ for a sufficiently large k-bit prime p. In [5] one can find another algorithm together with a heuristic argument that it works already for $\ell > (1/3+\varepsilon)k$, however it seems quite difficult to give a rigorous analysis of this algorithm, which can be a serious drawback in various cryptographic applications of ModInv-HNP (see [5] for some possible applications).

Throughout the paper we use $\log z$ to denote the binary logarithm of z.

1.2. Lattices and SVP problem

We recall that a lattice $\mathcal{L}$ is a set of all integer linear combinations of the form

$$\mathcal{L} = \left\{ \sum_{i=1}^{r} n_i \mathbf{b}_i \mid n_i \in \mathbb{Z} \right\},$$

for r linearly independent real vectors $\mathbf{b}_1, \ldots, \mathbf{b}_r$ in the s dimensional Euclidean space $\mathbb{R}^s$ (note that $r \leq s$). The set $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ is said to be a basis of $\mathcal{L}$.

One of the most fundamental problems in this area is the γ-*shortest vector problem*, γ-**SVP**: given a real $\gamma \geq 1$ and a basis of a lattice $\mathcal{L}$ in $\mathbb{R}^s$, find a nonzero vector $\mathbf{u} \in \mathcal{L}$, with the Euclidean norm $\|\mathbf{u}\|$ no more than γ times larger than the Euclidean norm of the shortest nonzero vector in $\mathcal{L}$, that is,

$$\|\mathbf{u}\| \leq \gamma \min\{\|\mathbf{w}\| \ : \ \mathbf{w} \in \mathcal{L}, \ \mathbf{w} \neq 0\}.$$

For $\gamma = 1$ this problem is known as simply **SVP**, we refer to [10–14] for the state of art and also surveys of previous results concerning different algorithms for γ-**SVP**.

1.3. Main results

We assume that we have access to a γ-SVP algorithm.

Theorem 1. Assume that for a prime number p we are given $n+1$ pairs

$$\left(t_i, \mathrm{MSB}_{\ell,p}\left(\frac{1}{\alpha + t_i}\right)\right), \qquad i = 1, \ldots, n+1,$$

with

$$(t_1, \ldots, t_{n+1}) \in (\mathbb{F}_p \setminus \{-\alpha\})^{n+1}$$

chosen uniformly at random. Then $\alpha \in \mathbb{F}_p$ can be recovered in deterministic polynomial time and a single call to a γ-SVP algorithm on a $(2n+2)$-dimensional lattice with polynomially bounded basis, except with probability

$$P \leq \frac{2^{n+1}(4h\Delta + 1)^{3n+1}}{(p-1)^n} + \frac{4(4h\Delta+1)^3}{p-1}$$

over the choices of $t_1, \ldots, t_{n+1}$, when it either returns no answer or returns a wrong answer, where

$$h = \gamma\sqrt{2n+2} \qquad \text{and} \qquad \Delta = \left\lceil p/2^{\ell+1} \right\rceil.$$

Theorem 1 implies that for almost all evaluation points, ModInv-HNP can be solved in deterministic polynomial time if $\ell > (2/3 + \varepsilon)k$ where k is the bit length of p (for any constant $\varepsilon > 0$). The following corollary gives a more precise statement, in two variants, using two different SVP approximation algorithms. Although the run-time is polynomial in k for any constant ε in both cases, the dependance on ε is different, and allows trading off a larger run-time for a smaller minimum allowed value of k.

Corollary 1. *Fix ε and δ with $0 < \delta < \varepsilon < 1$. Let*

$$n_0 = \left\lceil \frac{2}{9\varepsilon} \right\rceil,$$

let p be a k-bit prime and let $\ell > (2/3 + \varepsilon)k$. There exist deterministic algorithms $\mathcal{A}_1$ and $\mathcal{A}_2$ such that given $n_0 + 1$ pairs

$$\left(t_i, \mathrm{MSB}_{\ell,p}\left(\frac{1}{\alpha + t_i}\right)\right), \qquad i = 1, \ldots, n_0+1,$$

with

$$(t_1, \ldots, t_{n_0+1}) \in (\mathbb{F}_p \setminus \{-\alpha\})^{n_0+1}$$

chosen uniformly at random, for $k \ge k_\nu$ the algorithm $\mathcal{A}_\nu$ runs in time T_ν, $\nu = 1, 2$, and recovers $\alpha \in \mathbb{F}_p$ correctly with probability at least $1-p^{-\delta}$ over the choices of $t_1, \ldots, t_{n_0+1}$, where

$$k_1 = \lceil c_1 \varepsilon^{-1} \log \varepsilon^{-1} \rceil \qquad \text{and} \qquad k_2 = \left\lceil c_2 \varepsilon^{-2} \frac{(\log\log \varepsilon^{-1})^2}{\log \varepsilon^{-1}} \right\rceil,$$

for some absolute effectively computable constants c_1 and c_2, and

$$T_1 = (2^{\varepsilon^{-1}} k)^{O(1)} \qquad \text{and} \qquad T_2 = (\varepsilon^{-1} k)^{O(1)}.$$

Proof. Plugging $\Delta = 2^{k-\ell+O(1)}$ and $p = 2^{k+O(1)}$ in the first term of the error probability bound in Theorem 1, and using that $\ell > (2/3+\varepsilon)k$ we see that

$$\begin{aligned}\frac{2^{n+1}(4h\Delta+1)^{3n+1}}{(p-1)^n} &= h^{3n+1} 2^{(3n+1)((k-\ell)+O(1))-(k+O(1))n} \\ &\le h^{3n+1} 2^{(3n+1)(k/3-k\varepsilon+O(1))-(k+O(1))n} \\ &= 2^{(3n+1)\log h + k/3 - k(3n+1)\varepsilon + O(n)}.\end{aligned}$$

Now a straightforward computation shows that the first term is upper bounded by $p^{-\delta}/2$ if the condition

$$(3n+1)\cdot(\varepsilon - k^{-1}(\log h + C_0)) \ge 1/3 + \delta \tag{1}$$

is satisfied (with some sufficiently large absolute constant C_0). Assuming

$$k \ge (2\log h + C_1)\cdot \varepsilon^{-1} \tag{2}$$

for a sufficiently large absolute constant C_1, and using that $\delta < \varepsilon$, we see that $n \ge n_0$ implies (1).

Furthermore, another straightforward computation shows that the second term in the error probability of Theorem 1 is bounded by $p^{-\delta}/2$ if

$$k \ge (3\log h + C_2)\cdot(3\varepsilon - \delta)^{-1}.$$

(for a suitable absolute constant C_2). It is easy to see that this condition on k is implied by (1), provided that C_0 is large enough, using $\delta < \varepsilon$.

For $\mathcal{A}_1$, we apply the 1-SVP algorithm of [11] to a lattice of dimension $s = 2n_0 + 2$, which gives $h = \sqrt{2n_0+2}$.

For $\mathcal{A}_2$, we use the $2^{O(s(\log\log s)^2/\log s)}$-SVP algorithm of Schnorr [15] for $s = 2n_0+2$, which gives $h = 2^{n_0+1}\sqrt{2n_0+2}$. Recalling the definition of n_0, the stated bounds on k follow.

This completes the proof. □

Note that a trivial information theoretic lower bound on the number n_0+1 of pairs

$$\left(t_i, \mathrm{MSB}_{\ell,p}\left(\frac{1}{\alpha+t_i}\right)\right), \qquad i = 1, \ldots, n_0,$$

needed to recover α is

$$n_0 \ge \frac{k-1}{\ell} - 1,$$

since α has entropy at least $k-1$ bits, and each pair provides at most ℓ bits of information on α. Hence for the parameter choice $\ell > (2/3+\varepsilon)k$, our algorithm works with n_0 within a constant factor of the lower bound, with a constant that varies inversely with ε.

2. Proof of Theorem 1

2.1. Algorithm

We assume that

$$2h\Delta \le 6h\Delta^2 < p \tag{3}$$

since otherwise the result is trivial (as the claimed bound on the probability exceeds 1).

Assume we are given $n+1$ pairs of integers (t_i, u_i) with

$$u_i \equiv \frac{1}{\alpha + t_i} + e_i \pmod p \tag{4}$$

for some integers e_i with $|e_i| \le \Delta$, $i = 1, \ldots, n+1$.

Rewriting these congruences as

$$\alpha \equiv \frac{1}{u_i - e_i} - t_i \pmod p,$$

and eliminating α, we obtain

$$\frac{1}{u_1 - e_1} - t_1 \equiv \frac{1}{u_i - e_i} - t_i \pmod p, \qquad i = 2, \ldots, n+1,$$

which in turn implies

$$u_i - e_i - u_1 + e_1 \equiv (t_1 - t_i)(u_1 - e_1)(u_i - e_i) \pmod p, \tag{5}$$

for $i = 2, \ldots, n+1$.

Denoting

$$\begin{aligned} A_i &\equiv (t_1 - t_i)u_1u_i + u_1 - u_i \pmod p, \\ B_{1,i} &\equiv -(t_1 - t_i)u_i - 1 \pmod p, \\ B_{i,i} &\equiv -(t_1 - t_i)u_1 + 1 \pmod p, \\ C_i &\equiv t_1 - t_i \pmod p, \end{aligned}$$

we write (5) as

$$A_i + B_{1,i}e_1 + B_{i,i}e_i + C_ie_1e_i \equiv 0 \pmod p, \qquad i = 2, \ldots, n+1.$$

We now rescale the coefficients as

$$\begin{aligned} a_i \equiv A_i\Delta^{-2} \pmod p, \quad & b_{1,i} \equiv B_{1,i}\Delta^{-1} \pmod p, \\ b_{i,i} \equiv B_{i,i}\Delta^{-1} \pmod p, \quad & c_i \equiv C_i \pmod p, \end{aligned}$$

for $i = 2, \ldots, n+1$, and notice that the vector

$$\mathbf{e} = \left(\Delta^2, \Delta e_1, \ldots, \Delta e_{n+1}, e_1e_2, \ldots, e_1e_{n+1}\right)$$

belongs to the lattice $\mathcal{L}$ consisting of solutions

$$\mathbf{x} = (x_0, x_1, \ldots, x_{n+1}, x_{1,2}, \ldots, x_{1,n+1}) \in \mathbb{Z}^{2n+2}$$

of the congruences

$$\begin{aligned} a_ix_0 + b_{1,i}x_1 + b_{i,i}x_i + c_ix_{1,i} &\equiv 0 \pmod p, \qquad i = 2, \ldots, n+1. \\ x_0 &\equiv 0 \pmod{\Delta^2}, \\ x_j &\equiv 0 \pmod \Delta, \qquad j = 1, \ldots, n+1. \end{aligned} \tag{6}$$

We note that a $(2n+2) \times (2n+2)$ integral basis matrix

$$M = (M_{i,j})_{i,j=0}^{2n+1}$$

whose rows generate $\mathcal{L}$ can be constructed efficiently as follows:

The first two columns are defined as follows: $M_{0,0} = \Delta^2$ and $M_{i,0} = 0$ for $i = 1, \ldots, 2n+1$ (imposing the relation $x_0 \equiv 0 \pmod{\Delta^2}$), and $M_{1,1} = \Delta$ and $M_{i,1} = 0$ for $j \neq 1$ (relation $x_1 \equiv 0 \pmod{\Delta}$).

For $j \in \{2, \ldots, n+1\}$, there are two possible cases:

- If $C_j \not\equiv 0 \pmod{p}$, then C_j is invertible modulo p, and we set $M_{j,j} = \Delta$ and $M_{i,j} = 0$ for $i \neq j$ (we recall the relation $x_j \equiv 0 \pmod{\Delta}$) and $M_{0,n+j} \equiv -C_j^{-1} \cdot A_j \pmod{p}$, $M_{1,n+j} \equiv -C_j^{-1} \cdot B_{1,j} \pmod{p}$, $M_{j,n+j} \equiv -C_j^{-1} \cdot B_{j,j} \pmod{p}$, $M_{n+j,n+j} = p$, and $M_{i,n+j} = 0$ for $i \notin \{0, 1, j, n+j\}$ (we recall the relation $x_{1,i} \equiv -c_j^{-1} \cdot (a_j x_0 + b_{1,j} x_1 + b_{j,j} x_j) \pmod{p}$).
- If $C_j \equiv 0 \pmod{p}$, then the relations $x_j \equiv 0 \pmod{\Delta}$ and $x_j \equiv -a_j x_0 - b_{1,j} x_1 \pmod{p}$ hold. Since p and Δ are coprime, the latter two congruences are equivalent to the single congruence $x_j \equiv \Delta \cdot \lfloor \Delta^{-1} \rfloor_p \cdot (-a_j x_0 - b_{1,j} x_1) \pmod{p\Delta}$. Consequently, we set $M_{0,j} \equiv -\Delta^2 \cdot \lfloor \Delta^{-1} \rfloor_p \cdot A_j \pmod{p\Delta}$, $M_{1,j} \equiv -\Delta^2 \cdot \lfloor \Delta^{-1} \rfloor_p \cdot B_{1,j} \pmod{p\Delta}$, $M_{j,j} = p\Delta$ and $M_{i,j} = 0$ for $i \notin \{0, 1, j\}$ (relation $x_j \equiv \Delta \cdot \lfloor \Delta^{-1} \rfloor_p \cdot (-a_j x_0 - b_{1,j} x_1) \pmod{p\Delta}$), and $M_{n+j,n+j} = 1$ and $M_{i,n+j} = 0$ for $i \neq n+j$ (we recall the relation $x_{1,n+j} \in \mathbb{Z}$).

Clearly the Euclidean norm $\|\mathbf{e}\|$ of $\mathbf{e}$ satisfies the inequality

$$\|\mathbf{e}\| \leq \left(\Delta^4 + \cdots + \Delta^4\right)^{1/2} \leq \sqrt{2n+2}\Delta^2.$$

We run the γ-SVP algorithm on the lattice $\mathcal{L}$. Let

$$\mathbf{f} = (\Delta^2 f_0, \Delta f_1, \ldots, \Delta f_{n+1}, f_{1,2}, \ldots, f_{1,n+1}) \in \mathcal{L},$$

where $f_0, \ldots, f_{n+1}, f_{1,2}, \ldots, f_{1,n+1} \in \mathbb{Z}$, be the returned approximation to the shortest nonzero vector in $\mathcal{L}$. So

$$\|\mathbf{f}\| \leq \gamma \cdot \|\mathbf{e}\| \leq \gamma\sqrt{2n+2} \cdot \Delta^2 = h \cdot \Delta^2.$$

We have

$$\begin{aligned} |f_0| &\leq \|\mathbf{f}\|\Delta^{-2} \leq h, \\ |f_i| &\leq \|\mathbf{f}\|\Delta^{-1} \leq h \cdot \Delta, \qquad i = 1, \ldots, n+1, \\ |f_{1,i}| &\leq \|\mathbf{f}\| \leq h \cdot \Delta^2, \qquad i = 2, \ldots, n+1. \end{aligned} \tag{7}$$

We now consider the vector

$$\mathbf{d} = (0, \Delta d_1, \ldots, \Delta d_{n+1}, d_{1,2}, \ldots, d_{1,n+1}) = f_0 \mathbf{e} - \mathbf{f},$$

where

$$d_i = f_0 e_i - f_i, \qquad i = 1, \ldots, n+1,$$

and

$$d_{1,i} = f_0 e_1 e_i - f_{1,i} \qquad i = 2, \ldots, n+1.$$

Observe that if $d_i = 0$ for some $i \in \{1, \ldots, n+1\}$ and also $f_0 \neq 0$, then we can compute $e_i = f_i / f_0$. To decide which of the integral ratios f_i / f_0 is indeed equal to e_i, we perform the following "consistency check":

- We form the set $\mathcal{I} \subseteq \{1, \ldots, n+1\}$ of $i = 1, \ldots, n+1$ with integral values of f_i/f_0 and $u_i \not\equiv f_i/f_0 \pmod p$. For every $i \in \mathcal{I}$ we compute
$$\beta_i \equiv \frac{1}{u_i - f_i/f_0} - t_i \pmod p, \qquad 0 \le \beta_i < p,$$
and also define h_{ij}, $j = 1, \ldots, n+1$, by the conditions
$$h_{ij} \equiv \frac{1}{\beta_i + t_j} - u_j \pmod p, \qquad |h_{ij}| < p/2,$$
(again discarding the values of i for which at least one inversion fails).
- We now choose the smallest $i_0 \in \mathcal{I}$ for which for all $j = 1, \ldots, n+1$, the values $h_{i_0 j}$ exist and satisfy $|h_{i_0 j}| \le \Delta$, and return β_{i_0}. Otherwise we return failure.

2.2. Necessary conditions for failure

Let us define the quantities
$$E_i = d_{1,i} - d_1 e_i - d_i e_1, \qquad i = 2, \ldots, n+1. \tag{8}$$

Let us define the "bad" events
$$\begin{aligned} \mathcal{E}_1 :&\quad d_i \ne 0 \text{ or } E_i \ne 0, \quad \text{for every } i = 2, \ldots, n+1,\\ \mathcal{E}_2 :&\quad f_0 = 0 \text{ and } \mathcal{E}_1 \text{ does not hold},\\ \mathcal{E}_3 :&\quad \alpha \ne \beta_{i_0} \text{ and } \mathcal{E}_1 \text{ and } \mathcal{E}_2 \text{ do not hold}.\end{aligned}$$

Note that we ignore the case of $d_1 \ne 0$ as below we shown in Section 2.4 that $\mathcal{E}_2$ impllies $d_1 = 0$.

As observed above, our algorithm succeeds (that is, $\beta_{i_0} = \alpha$) if $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ do not occur.

We now upper bound the probability of these bad events over the choice of $t_1, \ldots, t_{n+1}$, chosen uniformly at random from $\mathbb{F}_p \setminus \{-\alpha\}$.

We first derive some useful relations satisfied by the difference vector $\mathbf{d}$. Using the first n congruences in (6), we find that
$$b_{1,i}\Delta d_1 + b_{i,i}\Delta d_i + c_i d_{1,i} \equiv 0 \pmod p. \tag{9}$$

Note that for $i = 1, \ldots, n+1$ we have
$$|d_i| = |f_0 e_i - f_i| \le h|e_i| + |f_i| \le 2h \cdot \Delta \tag{10}$$
and also for $i = 2, \ldots, n+1$,
$$|d_{1,i}| = |f_0 e_1 e_i - f_{1,i}| \le h|e_1 e_i| + |f_{1,i}| \le 2h \cdot \Delta^2. \tag{11}$$

We see that (9) implies
$$B_{1,i} d_1 + B_{i,i} d_i + C_i d_{1,i} \equiv 0 \pmod p, \qquad i = 2, \ldots, n+1. \tag{12}$$

Recalling the definition of $B_{1,i}, B_{i,i}, C_i$, we find that
$$-d_1\left((t_1 - t_i)u_i + 1\right) + d_i\left(-(t_1 - t_i)u_1 + 1\right) + d_{1,i}(t_1 - t_i) \equiv 0 \pmod p,$$
or
$$(t_1 - t_i)\left(-d_1 u_i - d_i u_1 + d_{1,i}\right) \equiv d_1 - d_i \pmod p.$$

Finally, using (4) we derive a quadratic congruence in t_i:

$$U_i \cdot t_i^2 + V_i \cdot t_i + W_i \equiv 0 \pmod p, \qquad i = 2, \ldots, n+1, \tag{13}$$

where

$$\begin{aligned} U_i &\equiv \frac{d_i}{\alpha + t_1} - E_i \pmod p, \\ V_i &\equiv (t_1 - \alpha) \cdot \left(E_i - \frac{d_i}{\alpha + t_1} \right) + d_i \pmod p, \\ W_i &\equiv \alpha \cdot (d_i - d_1) + t_1 \cdot \left(\frac{-d_i \alpha}{\alpha + t_1} - d_1 + \alpha E_i \right) \pmod p, \end{aligned} \tag{14}$$

and E_i is given by (8), $i = 2, \ldots, n+1$.

2.3. Estimating the probability of $\mathcal{E}_1$

Assume that $\mathcal{E}_1$ holds. Let us fix some values of t_1, d_i for $i = 1, \ldots, n+1$, and $d_{1,i}$ for $i = 2, \ldots, n+1$. We now consider, the number of n-tuples

$$(t_2, \ldots, t_{n+1}) \in (\mathbb{F}_p \setminus \{-\alpha\})^n$$

satisfying (13).

We claim that for every $i = 2, \ldots, n+1$, the left hand side of (13) is a non-constant polynomial of degree at most 2 in t_i and hence has at most 2 solutions for t_i. Thus we have at most 2^n such n-tuples.

Indeed, by our assumption, we know that either $d_i \neq 0$ or $E_i \neq 0$ holds.

In the case $d_i \neq 0$, using the inequality $|d_i| \leq 2h \cdot \Delta < p$, see (3), we have $d_i \not\equiv 0 \pmod p$. There are two subcases to consider:

- If $E_i \equiv 0 \pmod p$, then (14) shows that $U_i \not\equiv 0 \pmod p$.
- If $E_i \not\equiv 0 \pmod p$ then $U_i \equiv 0 \pmod p$ if and only if we have $\alpha + t_1 \equiv d_i/E_i \pmod p$, which implies that $V_i \equiv d_i \pmod p$ and hence $V_i \not\equiv 0 \pmod p$.

So the claim holds if $d_i \neq 0$.

In the case $d_i = 0$ and $E_i \neq 0$, from (11) we have that $|E_i| \leq 6h\Delta^2 < p$, so, recalling (3), we see that $E_i \not\equiv 0 \pmod p$, and then (14) shows that $U_i \not\equiv 0 \pmod p$, as claimed.

Now, we see from (10) that the tuple $(d_1, \ldots, d_{n+1})$ can take at most $(4h\Delta + 1)^{n+1}$ possible values. Furthermore, using the inequality $|E_i| \leq 6h\Delta^2$, we also see that the tuple $(E_2, \ldots, E_{n+1})$ takes at most $(12h\Delta^2 + 1)^n$ possible values. Since t_1 can take $p - 1$ possible values and $(t_2, \ldots, t_{n+1})$ at most 2^n possible values, we conclude that there are at most

$$2^n (4h\Delta + 1)^{n+1} (12h\Delta^2 + 1)^n (p-1) < 2^n (4h\Delta + 1)^{3n+1} (p-1)$$

tuples $(t_1, \ldots, t_{n+1})$ for which the bad event $\mathcal{E}_1$ happens. So the probability of $\mathcal{E}_1$ is at most

$$\Pr[\mathcal{E}_1] \leq \frac{2^n (4h\Delta + 1)^{3n+1}}{(p-1)^n}. \tag{15}$$

2.4. Estimating the probability of $\mathcal{E}_2$

We first claim that $\mathcal{E}_2$ implies $d_1 = 0$. Indeed, $\neg\mathcal{E}_1$ means that there exists $j \in \{2, \ldots, n+1\}$ such that $d_j = 0$ and $E_j = 0$. Then from (14), we see that $U_j = V_j = 0$ and $W_j = -(\alpha + t_1) \cdot d_1$. Hence (13) implies that $W_j \equiv 0 \pmod p$, which leads to $d_1 \equiv 0 \bmod p$ (since $\alpha + t_1 \not\equiv 0 \pmod p$) and hence $d_1 = 0$ (since $|d_1| < 2h\Delta < p$).

Now, let us consider the set

$$S = \{i \in \{2, \ldots, n+1\} : E_i = 0 \text{ and } d_i = 0\},$$

and denote its size by k. We claim that $k < n$. Indeed, from (8), we see that for each $i \in S$ we have $E_i = d_{1,i}$. Therefore, if $k = n$ then we have $d_i = 0$ for $i = 1, \ldots, n+1$ and $d_{1,i} = 0$ for $i = 2, \ldots, n+1$, which implies (since $f_0 = 0$) that $\mathbf{f} = \mathbf{0}$, a contradiction. Thus we must have $k < n$.

Now, let us fix t_1, d_i for $i = 2, \ldots, n+1$, and $d_{1,i}$ for $i = 2, \ldots, n+1$ and consider, for $i = 2, \ldots, n+1$ the number of solutions for t_i satisfying (13). If $i \in S$, then the left hand side of (13) is the zero polynomial so t_i has p possible values. If $i \notin S$, we have $d_i \neq 0$ or $E_i \neq 0$ so (as shown in the analysis of $\mathcal{E}_1$ above) the left hand side of (13) is a non-constant polynomial of degree at most 2 and hence there are at most 2 possible values for t_i. Overall there are at most $2^{n-k}(p-1)^k$ solutions for $(t_2, \ldots, t_{n+1})$. There are $p-1$ possible values for t_1. Furthermore, as before we see that there are at most $(12h\Delta^2+1)^{n-k}$ possible values for E_i and $(4h\Delta+1)^{n-k}$ possible values for d_i with $i \notin S$. So overall, there are at most

$$\begin{aligned} &2^{n-k}(4h\Delta+1)^{n-k}(12h\Delta^2+1)^{n-k}(p-1)^{k+1} \\ &\qquad\qquad < 2^{n-k}(4h\Delta+1)^{3(n-k)}(p-1)^{k+1} \end{aligned}$$

tuples $(t_1, \ldots, t_{n+1})$ for which the bad event $\mathcal{E}_2$ happens. So, since $k < n$, the probability of $\mathcal{E}_2$ is at most

$$\Pr[\mathcal{E}_2] \leq \sum_{k=0}^{n-1} \left(\frac{2(4h\Delta+1)^3}{p-1} \right)^{n-k} \leq \sum_{r=1}^{\infty} \left(\frac{2(4h\Delta+1)^3}{p-1} \right)^{r}. \tag{16}$$

We see that

$$\Pr[\mathcal{E}_2] \leq \frac{4(4h\Delta+1)^3}{p-1}. \tag{17}$$

Indeed, for $2(4h\Delta+1)^3/(p-1) \geq 1/2$ it is obvious as we always have $\Pr[\mathcal{E}_2] \leq 1$; otherwise it follows from (16).

2.5. Estimating the probability of $\mathcal{E}_3$

If $\mathcal{E}_3$ holds, then we have that β_{i_0} and $\alpha \neq \beta_{i_0}$ satisfy the relations

$$\frac{1}{\beta_{i_0} + t_j} - u_j \equiv h_{i_0 j} \pmod p$$

and

$$\frac{1}{\alpha + t_j} - u_j \equiv e_j \pmod p,$$

for $j = 1, \dots, n+1$. Subtracting the last two relations and multiplying by $(\alpha+t_j)\cdot(\beta_{i_0}+t_j)$ we obtain the relation

$$\alpha - \beta_{i_0} \equiv (h_{i_0 j} - e_j)(\alpha + t_j)(\beta_{i_0} + t_j) \mod p,$$

with $|h_{i_0 j}| \le \Delta$, $j = 1, \dots, n+1$. Clearly for every fixed β_{i_0} and $h_{i_0 j}$ there are at most two possible values of t_j. Since there are $p-1$ possibilities of β_{i_0} and at most $2\Delta + 1$ possibilities for every $h_{i_0 j}$, $j = 1, \dots, n+1$, as before we conclude that

$$\Pr[\mathcal{E}_3] \le \frac{(2\Delta+1)^n(p-1)}{(p-1)^{n+1}} = \frac{(2\Delta+1)^n}{(p-1)^n} \le \Pr[\mathcal{E}_1]$$

which together with (15) and (17) concludes the proof.

3. Remarks

We note that a slightly more careful analysis of the event $\mathcal{E}_2$ in the proof of Theorem 1 allows to show that $k = n-1$ with probability $O(\Delta^2/p)$ which improves the second term in the probability estimate of Theorem 1. This however does not change the 2/3-threshold in Corollary 1. It remains a challenging open problem to get a rigorous version of the other (presumably more powerful) algorithm of [5] (or its appropriate modification) which can potentially lead to the replacing 2/3 with 1/3. This algorithm is based on the ideas of Coppersmith [8,9]. However the rigorous analysis of this approach seems to be much more difficult which we pose an open question.

It is also interesting to check whether the recently emerged approach of Akavia [1] can be applied to ModInv-HNP.

More generally, it is certainly interesting to study a general problem of recovering of an unknown rational function $\psi(X) \in \mathbb{F}_p(X)$ from a sequence of k pairs $(t_i, \mathrm{MSB}_{\ell,p}(\psi(t_i)))$, $i = 1, \dots, k$.

Finally, it appears that ModInv-HNP in the case when the modulus p is also hidden is a much more difficult problem to which no feasible approaches are known at the moment. Thus this could be a very promising cryptographic primitive.

Acknowledgements

This paper was initiated during a very enjoyable visit of I. S. and R. S. to School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, whose hospitality, support and stimulating research atmosphere are gratefully appreciated.

This research was supported by NRF Grant CRP2-2007-03, Singapore (for all co-authors), by ARC Grant DP1092835, Australia (for I. S.) and by ARC Grant DP0987734, Australia (for R. S.)

References

[1] A. Akavia, 'Solving hidden number problem with one bit oracle and advice', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **5677** (2010), 337–354.

[2] S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, 'Predicting the inversive generator', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **2898** (2003), 264–275.

[3] S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, 'Predicting nonlinear pseudorandom number generators', *Math. Comp.*, **74** (2005), 1471–1494.

[4] S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, 'Reconstructing noisy polynomial evaluation in residue rings', *J. of Algorithms*, **61** (2006), 47–90.

[5] D. Boneh, S. Halevi and N. A. Howgrave-Graham, 'The modular inversion hidden number problem', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **2248** (2001), 36–51.

[6] D. Boneh and R. Venkatesan, 'Hardness of computing the most significant bits of secret keys in Diffie–Hellman and related schemes', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **1109** (1996), 129–142.

[7] D. Boneh and R. Venkatesan, 'Rounding in lattices and its cryptographic applications', *Proc. 8th Annual ACM-SIAM Symp. on Discr. Algorithms*, SIAM, 1997, 675–681.

[8] D. Coppersmith, 'Small solutions to polynomial equations, and low exponent RSA vulnerabilities', *J. Cryptology*, **10** (1997), 233–260.

[9] D. Coppersmith, 'Small solutions of small degree polynomials', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **2146** (2001), 20–31.

[10] N. Gama and P. Q. Nguyen, 'Finding short lattice vectors within Mordell's inequality', *Proc. 40 ACM Symp. on Theory of Comp.*, ACM, 2008, 207–216.

[11] D. Micciancio and P. Voulgaris, 'A deterministic single exponential time algorithm for most lattice problems based on Voronoi cell computations', *Electronic Colloq. on Comp. Compl.*, **14** (2010), 1–14 (available from `http://www.eccc.uni-trier.de/report/2010/014/`).

[12] P. Q. Nguyen and D. Stehlé, 'An LLL algorithm with quadratic complexity', *SIAM J. Comput.*, **39** (2009), 874–903.

[13] A. Novocin, D. Stehlé and G. Villard, 'An LLL-reduction algorithm with quasi-linear time complexity', *Proc. 43rd ACM Symp. Theory of Comp.* (to appear).

[14] X. Pujol and D. Stehlé, 'Solving the shortest lattice vector problem in time $2^{2.465n}$', *Cryptology ePrint Archive*, Report 2009/605, 2009, (available from `http://eprint.iacr.org/2009/605`).

[15] C. P. Schnorr, 'A hierarchy of polynomial time basis reduction algorithms', *Theor. Comp. Sci.*, **53** (1987), 201–224.

[16] I. E. Shparlinski, 'Playing "Hide-and-Seek" with numbers: The hidden number problem, lattices and exponential sums', *Proc. Symp. in Appl. Math.*, Amer. Math. Soc., Providence, RI, **62** (2005), 153–177.