A HIDDEN NUMBER PROBLEM IN SMALL SUBGROUPS

IGOR SHPARLINSKI AND ARNE WINTERHOF

Abstract. Boneh and Venkatesan have proposed a polynomial time algorithm for recovering a hidden element $\alpha \in \mathbb{F}_p$, where $p$ is prime, from rather short strings of the most significant bits of the residue of $\alpha t$ modulo $p$ for several randomly chosen $t \in \mathbb{F}_p$. González Vasco and the first author have recently extended this result to subgroups of $\mathbb{F}_p^*$ of order at least $p^{1/3+\varepsilon}$ for all $p$ and to subgroups of order at least $p^\varepsilon$ for almost all $p$. Here we introduce a new modification in the scheme which amplifies the uniformity of distribution of the multipliers $t$ and thus extend this result to subgroups of order at least $(\log p)/(\log \log p)^{1-\varepsilon}$ for all primes $p$. As in the above works, we give applications of our result to the bit security of the Diffie–Hellman secret key starting with subgroups of very small size, thus including all cryptographically interesting subgroups.

1. Introduction

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 a real $\eta > 0$ and $t \in \mathbb{F}_p$ we denote by $\text{MSB}_\eta(t)$ any integer which satisfies the inequality

$$|t - \text{MSB}_\eta(t)| < p^{2^{-\eta-1}}.$$ 

Roughly speaking, $\text{MSB}_\eta(t)$ is an integer having about $\eta$ most significant bits as $t$ (taking into account our convention about the elements of $\mathbb{F}_p$). However, this definition is more flexible and better suited to our purposes. In particular we remark that $\eta$ in the inequality (1) need not be an integer.

Given a subgroup $G \subseteq \mathbb{F}_p^*$, we consider the following hidden number problem over $G$.

Recover a number $\alpha \in \mathbb{F}_p$ such that for $d$ elements $t_1, \ldots, t_d \in G$, chosen independently and uniformly at random from $G$, we are given $d$ pairs

$$(t_h, \text{MSB}_\eta(\alpha t_h)), \quad h = 1, \ldots, d,$$

for some $\eta > 0$.

For $G = \mathbb{F}_p^*$ this problem has been introduced and studied by Boneh and Venkatesan [1]. In [1] a polynomial time algorithm is designed that recovers $\alpha$ for some
\begin{align*}
\eta \sim (\log p)^{1/2} \text{ and } d = O((\log^{1/2} p)). \text{ The algorithm of } [1] \text{ has been extended in several directions. In particular, in } [8] \text{ it is generalised to all sufficiently large subgroups } 
\mathcal{G} \subseteq \mathbb{F}_p^*. \text{ This and other generalisations have led to a number of cryptographic applications, see } [22, 23, 24]. \text{ Using bounds of exponential sums from } [9, 12], \text{ it has been shown that the algorithm of } [1] \text{ works for subgroups } 
\mathcal{G} \subseteq \mathbb{F}_p^* \text{ of order } # \mathcal{G} \geq p^{\nu + \varepsilon} \text{ where for any } \varepsilon > 0 \text{ and sufficiently large } Q \text{ one can take } 
\begin{itemize}
\item \nu = 1/3 \text{ for all primes } p \in [Q, 2Q], 
\item \nu = 0 \text{ for all primes } p \in [Q, 2Q], \text{ except for an exceptional set of size at most } Q^{6/6 + \varepsilon}. 
\end{itemize}
\end{align*}

Using a recent improvement of [3] of the bounds of exponential sums, one can obtain the same result with \nu = 0 \text{ for all primes } p \in [Q, 2Q]. \text{ However, in practical applications, one would probably choose a smaller subgroup of size about } \exp(c(\log p)^{1/3}(\log \log p)^{2/3}) \text{ for some constant } c > 0 \text{ in order to balance complexities of the attacks based on the number field sieve and those based on the Pollard rho method; see } [6, 15, 18, 19, 27].

Here we extend the algorithm of [1] to the case of almost arbitrary subgroups \mathcal{G} \subseteq \mathbb{F}_p^*. \text{ More precisely, our result applies to any subgroup } \mathcal{G} \subseteq \mathbb{F}_p^* \text{ of size } # \mathcal{G} \geq \log p/(\log \log p)^{1-\varepsilon}; \text{ thus, it includes all subgroups of cryptographically interesting sizes. As in [1], our method is based on lattice reduction algorithms, and it also makes use of exponential sums, though not in such a direct way as in [8]. Namely, we introduce certain new arguments allowing us to amplify the uniformity of distribution properties of small subgroups } \mathcal{G}. \text{ This allows us to use the bound of exponential sums from [10] with elements of } \mathcal{G}, \text{ which is very moderate in strength (and does not imply any uniformity of distribution properties of } \mathcal{G} \text{ which would be the crucial argument of the method of [8]). The bound of [10], however, has the very important advantage over the bounds of [3, 9, 11, 12] that it applies to subgroups of order } # \mathcal{G} \geq \log p/(\log \log p)^{1-\varepsilon}. \text{ It is interesting to note that our approach has links with the famous Waring problem which has been studied in number theory for several hundred years. In fact, the Waring problem in finite fields has been the main motivation of the bound of exponential sums of [10] which we use in this paper. For surveys of recent results on this problem, see [3, 10, 28]. As in [1, 8] we apply our algorithm for the hidden number problem to derive a bit security result for the Diffie–Hellman scheme.}

We hope that similar ideas can be used for several other variants of the hidden number problem and its applications; see [22, 23, 24]. \text{ It has also found some applications to a problem on the complexity of approximation of the permanent [25] and to noisy polynomial interpolation [26]. On the other hand, it should be remarked that our approach requires an increase in the number of } t \in \mathcal{G} \text{ for which MSB}_t(\alpha t) \text{ is requested (which remains polynomial nevertheless, in particular it never exceeds } (\log p)^4). \text{ Throughout this paper } \log x \text{ always denotes the binary logarithm of } x > 0 \text{ and the constants in the "O" symbols may occasionally, where obvious, depend on a small positive parameter } \varepsilon \text{ and are absolute otherwise.}

We always assume that } p \text{ is a prime number with } p \geq 5; \text{ thus, the expressions } \log \log p \text{ and } \log \log \log p \text{ are defined (and positive).}

\textbf{Acknowledgments.} \text{ The first author was supported in part by ARC grant DP0211459. The second author was supported in part by DSTA grant R-394-000-011-422, by FWF grant S8313, and by the Austrian Academy of Sciences.}
2. Exponential sums and distribution of short sums of elements of subgroups

For a complex $z$ we put $e_p(z) = \exp(2\pi i z/p)$.

Let $T = \#G$, $T/(p - 1)$, be the cardinality of a subgroup $G \subseteq \mathbb{F}_p^\times$. If we put $n = (p - 1)/T$, then each element $r \in G$ has exactly $n$ representations $r = x^n$ with $x \in \mathbb{F}_p^\times$. Therefore, for any $\lambda \in \mathbb{F}_p$,

$$\sum_{r \in G} e_p(\lambda r) = \frac{T}{p - 1} \sum_{x \in \mathbb{F}_p^\times} e_p(\lambda x^n).$$

Now by Theorem 1 of [10] we have the following bound; see also [4, 11, 12].

Lemma 1. For any $1 > \varepsilon > 0$ there exists a constant $c(\varepsilon) > 0$ such that for any subgroup $G \subseteq \mathbb{F}_p^\times$ of order

$$T \geq \frac{\log p}{(\log \log p)^{1 - \varepsilon}}$$

the bound

$$\max_{\gcd(\lambda, p) = 1} \left| \sum_{r \in G} e_p(\lambda r) \right| \leq T \left( 1 - \frac{c(\varepsilon)}{(\log p)^{1+\varepsilon}} \right)$$

holds.

For an integer $k \geq 1$, a subgroup $G \subseteq \mathbb{F}_p^\times$, and $t \in \mathbb{F}_p$, we denote by $N_k(G, t)$ the number of solutions of the congruence

$$r_1 + \cdots + r_k \equiv t \pmod{p}, \quad r_1, \ldots, r_k \in G.$$

Recalling the relation between the set of $n$th powers, where $n = (p - 1)/T$, we see that studying the above congruence is equivalent to studying the congruence

$$x_1^n + \cdots + x_k^n \equiv t \pmod{p}, \quad x_1, \ldots, x_k \in \mathbb{F}_p^\times.$$

The problem of finding the smallest possible value of $k$ for which the congruence (or, in more traditional settings, the corresponding equation over $\mathbb{Z}$) has a solution for any $t$ is known as the Waring problem. However, for our purposes just solvability is not enough. Rather we need an asymptotic formula for the number of solutions.

We show that Lemma 1 can be used to prove that for reasonably small $k$, $N_k(G, t)$ is close to its expected value $T^k/p$.

Lemma 2. For any $1 > \varepsilon > 0$ there exists a constant $C(\varepsilon) > 0$ such that for any subgroup $G \subseteq \mathbb{F}_p^\times$, of order

$$T \geq \frac{\log p}{(\log \log p)^{1-\varepsilon}}$$

the bound

$$\max_{t \in \mathbb{F}_p} \left| N_k(G, t) - \frac{T^k}{p} \right| \leq \frac{T^k}{p^2}$$

holds for any integer $k \geq C(\varepsilon)(\log p)^{2+\varepsilon}$.

Proof. The well-known identity (see for example [14, Chapter 5.1])

$$\sum_{\lambda=0}^{p-1} e_p(\lambda u) = \begin{cases} 0, & \text{if } u \not\equiv 0 \pmod{p}, \\ p, & \text{if } u \equiv 0 \pmod{p}, \end{cases}$$
implies that
\[
N_k(G,a) = \sum_{r_1,\ldots,r_k \in G} \frac{1}{p} \sum_{\lambda=0}^{p-1} e_p \left( \lambda (r_1 + \cdots + r_k - t) \right)
\]
\[
= \frac{1}{p} \sum_{r \in G} \sum_{\lambda=0}^{p-1} e_p (-\lambda t) \left( \sum_{\lambda=0}^{p-1} e_p (\lambda r) \right)^k.
\]
Separating the term \(T^k/p\), corresponding to \(\lambda = 0\), and applying Lemma 1 to other terms, we obtain
\[
\max_{t \in \mathbb{F}_p} \left| N_k(G,t) - \frac{T^k}{p} \right| \leq T^k \left( 1 - \frac{c(\epsilon)}{(\log p)^{1+\epsilon}} \right)^k
\]
and the desired result follows. 

3. Uniform distribution and hidden number problem

First of all we recall the classical definition of the discrepancy \(D(\Gamma)\) of an \(N\)-element sequence \(\Gamma = (\gamma_1,\ldots,\gamma_N)\) of elements of the interval \([0,1] \):
\[
D(\Gamma) = \sup_{J \subseteq [0,1]} \left| \frac{A(J,N)}{N} - |J| \right|
\]
where the supremum is extended over all subintervals \(J\) of \([0,1]\), \(|J|\) is the length of \(J\), and \(A(J,N)\) denotes the number of points \(\gamma_n\) in \(J\) for \(0 \leq n \leq N-1\). Informally speaking the discrepancy tells us how much the number of hits \(A(J,N)\) of a given interval \(J\) differs from its expected value \(|J|N\).

Now, following [16] we say that a finite sequence \(T\) of elements of \(\mathbb{F}_p\) is \(\Delta\)-homogeneously distributed modulo \(p\) if for any integer \(\lambda\) with \(\gcd(\lambda,p) = 1\), the discrepancy of the sequence of fractions \((\lambda t/p)_{t \in T}\) is at most \(\Delta\) (we certainly assume that \(\lambda t \in \mathbb{F}_p\) and thus is reduced modulo \(p\) and all these points belong to \([0,1]\)).

The following statement is a generalization of Theorem 1 of [1] and is given in [16] as Lemma 4.

Lemma 3. Let \(\omega > 0\) be an arbitrary absolute constant. For a prime \(p\), define
\[
\eta = \omega \sqrt{\frac{\log p \log \log p}{\log \log p}} \quad \text{and} \quad d = \left\lceil 3\eta^{-1} \log p \right\rceil.
\]
Let \(T\) be a \(2^{-\eta}\)-homogeneously distributed modulo \(p\) sequence of integer numbers. There exists a probabilistic polynomial-time algorithm \(A\) such that for any \(\alpha \in \mathbb{F}_p\) given \(2d\) integers
\[
t_h \quad \text{and} \quad u_h = \text{MSB}_\eta (\alpha t_h), \quad h = 1,\ldots,d,
\]
its output satisfies for sufficiently large \(p\)
\[
\Pr \{ A(p,t_1,\ldots,t_d; u_1,\ldots,u_d) = \alpha \} \geq 1 - \frac{1}{p},
\]
where the probability is taken over all \(t_1,\ldots,t_d\) chosen uniformly and independently at random from the elements of \(T\) and all coin tosses of the algorithm \(A\).
4. Hidden number problem in subgroups

Assume that for \( \alpha \in \mathbb{F}_p^* \) and a subgroup \( \mathcal{G} \subseteq \mathbb{F}_p^* \) of order \( T \), generated by \( g \in \mathbb{F}_p^* \), we are given an oracle \( \mathcal{HNP}_\mu \) such that for every \( x \in \{0, 1, \ldots, T - 1\} \), it returns \( \text{MSB}_\mu(xg^x) \).

**Theorem 4.** Let \( \vartheta > 0 \) be an arbitrary absolute constant and let

\[
\mu = \vartheta \left( \frac{\log p \log \log \log p}{\log \log p} \right)^{1/2}.
\]

There exists a polynomial time probabilistic algorithm which, for any \( 1 > \varepsilon > 0 \) and any \( g \in \mathbb{F}_p^* \) of multiplicative order

\[
T \geq \frac{\log p}{(\log \log p)^{1-\varepsilon}}.
\]

makes \( O(\mu^{-1}(\log p)^{3+\varepsilon}) \) calls of the oracle \( \mathcal{HNP}_\mu \) and then recovers \( \alpha \) with probability at least \( 1 + O(2^{-\mu/2}) \).

**Proof.** For an integer \( w \) we denote by \( \lfloor w \rfloor_p \) the remainder of \( w \) on division by \( p \). Take \( C(\varepsilon) \) from Lemma 2 and put

\[
k = \left\lfloor C(\varepsilon)(\log p)^{2+\varepsilon} \right\rfloor, \quad \eta = 2\mu/3, \quad d = \left\lceil 3\eta^{-1} \log p \right\rceil.
\]

Then by Lemma 2, the sequence

\[
T = (r_1 + \cdots + r_k \mid r_1, \ldots, r_k \in \mathcal{G})
\]

of \( k \)-sums of elements of \( \mathcal{G} \) is \( 2^{-\eta} \)-homogeneously distributed modulo \( p \). Now we call the oracle \( \mathcal{HNP}_\mu \) for \( dk \) uniformly and independently at random chosen

\[
r_{11}, \ldots, r_{1k}, \ldots, r_{d1}, \ldots, r_{dk} \in \mathcal{G},
\]

and we get integers \( u_{hj} \) with

\[
|\lfloor \alpha r_{hj} \rfloor_p - u_{hj}| < p/2^{\mu+1}, \quad h = 1, \ldots, d, \quad j = 1, \ldots, k.
\]

For \( h = 1, 2, \ldots, d \), we put

\[
v_h = \sum_{j=1}^k \lfloor \alpha r_{hj} \rfloor_p, \quad t_h = \left\lfloor \sum_{j=1}^k r_{hj} \right\rfloor_p, \quad u_h = \sum_{j=1}^k u_{hj},
\]

where we used addition over \( \mathbb{Z} \).

Note that for sufficiently large \( p \),

\[
|v_h - u_h| < kp/2^{\mu+1} \leq p/2^{\eta+1}.
\]

Next, we have

\[
v_h - u_h - \lfloor v_h \rfloor_p + \lfloor u_h \rfloor_p = \nu p \quad \text{with } \nu \in \{-1, 0, 1\}.
\]

If \( \nu = 1 \), then we have

\[
\lfloor v_h \rfloor_p - \lfloor u_h \rfloor_p + p = |v_h - u_h| < p/2^{\eta+1},
\]

which is only possible if \( |v_h| < p/2^{\eta+1} \). If \( \nu = -1 \), then we have

\[
\lfloor u_h \rfloor_p - \lfloor v_h \rfloor_p + p = |v_h - u_h| < p/2^{\eta+1},
\]
which is only possible if \( |v_h|_p > p - p/2^\eta + 1 \). If \( \nu = 0 \), then we have
\[
|\lfloor \alpha t_h \rfloor_p - \lfloor u_h \rfloor_p| = |\lfloor v_h \rfloor_p - \lfloor u_h \rfloor_p| = |v_h - u_h| < \frac{p}{2^\eta + 1}.
\]
By Lemma 2 the probability that \( p/2^\eta + 1 \leq |v_h|_p \leq p - p/2^\eta + 1 \) for all \( h = 1, \ldots, d \) is \( 1 + O(d2^{-\eta}) \). Now the algorithm of Lemma 3 yields the correct \( \alpha \) with probability at least \( 1 + O(d2^{-\eta} + p^{-1}) = 1 + O(2^{-\nu}/2) \) if \( p \) is sufficiently large. \( \square \)

5. Bit security of the Diffie–Hellman scheme

As in [18] we now apply Theorem 4 to derive a bit security result for the Diffie–Hellman scheme.

We assume that for a subgroup \( G \subseteq \mathbb{F}_p^\ast \), we are given an oracle \( \mathcal{DH}_p \), which, given the values of \( X = g^x \in \mathbb{F}_p \) and \( Y = g^y \in \mathbb{F}_p \), outputs the value of \( \text{MSB}_p (g^{xy}) \).

Repeating the same arguments of [8], we derive the following result.

**Theorem 5.** Let \( \vartheta > 0 \) be an arbitrary absolute constant and let
\[
\mu = \vartheta \left( \frac{\log p \log \log \log p}{\log \log p} \right)^{1/2}.
\]

There exists a polynomial time probabilistic algorithm which, for any \( 1 > \varepsilon > 0 \) and any \( g \in \mathbb{F}_p^\ast \) of multiplicative order \( T \geq 1 \) and which, for any pair \( (a, b) \in \{0, 1, \ldots, T - 1\}^2 \), given the values of \( A = g^a \in \mathbb{F}_p \) and \( B = g^b \in \mathbb{F}_p \), makes \( O(\mu^{-1}(\log p)^{3+\varepsilon}) \) calls of the oracle \( \mathcal{DH}_p \) and computes \( g^{ab} \) correctly with probability \( 1 + O(2^{-\nu/2} + T^{-\varepsilon/2}) \).

**Proof.** For \( T < (\log p)^{1+\varepsilon} \) the result is trivial (one can compute \( g^{ab} \) in polynomial time without any oracle at all).

We now assume that \( T \geq (\log p)^{1+\varepsilon} \). As in [8], given a pair \( (a, b) \in \{0, 1, \ldots, T - 1\}^2 \), let us select an integer \( s \in \{0, 1, \ldots, T - 1\} \) uniformly at random. We compute
\[
g_s = Bg^s;
\]
thus, \( g_s = g^{b+s} \).

Let \( \Delta = T^{\vartheta/(1+\varepsilon)} \). The probability that \( \gcd(b+s, T) \geq \Delta \) is at most \( \tau(T)/\Delta \), where \( \tau(T) \) is the number of positive integer divisors of \( T \). Indeed, for any divisor \( D \mid T \) with \( D \geq \Delta \), there are at most \( T/D \leq T/\Delta \) values of \( x \in \{0, 1, \ldots, T - 1\} \) with \( \gcd(x, T) = D \).

Using the bound \( \tau(T) = T^{o(1)} \) (see [17], Chapter 1, Theorem 5.2), we obtain that the probability of \( \gcd(b+s, T) \geq \Delta \) is at most \( T^{-\varepsilon/2} \).

In the opposite case, when \( \gcd(b+s, T) < \Delta \), the multiplicative order of \( g_s \) is
\[
T_s = \frac{T}{\gcd(b+s, T)} > \frac{\log p}{(\log \log p)^{1-\varepsilon}}.
\]
Put \( \alpha_s = g^{a(b+s)} \). Now we can call the oracle \( \mathcal{DH}_p \) with \( g^x A = g^{x+a} \) and \( g_s \) to evaluate
\[
\text{MSB}_p (g^{(a+x)(b+s)}) = \text{MSB}_p (\alpha_s g_x^s)
\]
for an integer \( x \) chosen uniformly at random in the set \( \{0, 1, \ldots, T - 1\} \). Because \( T \mid T \), the values of \( x \) are uniformly distributed modulo \( T_s \) as well; thus, Theorem 4 can be applied, producing the desired result. \( \square \)
A HIDDEN NUMBER PROBLEM IN SMALL SUBGROUPS

6. Remarks

It is important to note that the implicit constants in our estimates can be explicitly evaluated and thus our arguments can be implemented in an effective algorithm.

For cryptographic applications only the case of prime $T$ is really important. Under this condition the proof of Theorem 5 can be simplified (and actually leads to a slightly stronger result).

It seems very plausible that the same approach can be applied to the hidden number problem over extension fields which is related to proving bit security of the XTR and LUC protocols (see [13, 21]), as well as of some tripartite key exchange protocols on elliptic curves (see [7]). In particular, for this purpose one can probably use a very weak bound of [20] which, however, applies to very small subgroups.

As in [11] Theorem 4 can be applied to a number of other cryptographic protocols, for example to the ElGamal cryptosystem.

In [2] a nonuniform algorithm has been constructed which works with much smaller values $\eta \sim \log \log p$. This means that if the points $t_1, \ldots, t_d \in G$ are known in advance, one can design (in exponential time) a certain data structure, that now given $d$ values $\text{MSB}_\eta(\alpha t_h), h = 1, \ldots, d$, the hidden number $\alpha$ can be found in polynomial time. A slight modification of our approach can be used to extend the algorithm of [2] as well; see [25].

References


Department of Computing, Macquarie University, Sydney, New South Wales 2109, Australia
E-mail address: igor@ics.mq.edu.au

RICAM, Austrian Academy of Sciences, Altenbergerstrasse 69, 4040 Linz, Austria
E-mail address: arne.winterhof@oeaw.ac.at