A Parameterized Splitting System and its Application to the Discrete Logarithm Problem with Low Hamming Weight Product Exponents

Sungwook Kim and Jung Hee Cheon

Department of Mathematical Sciences and ISaC-RIM,
Seoul National University, Seoul, 151-747, Korea
{avell7,jhcheon}@snu.ac.kr

Abstract. A low Hamming weight product (LHWP) exponent is used to increase the efficiency of cryptosystems based on the discrete logarithm problem (DLP). In this paper, we introduce a new tool, called a *Parameterized Splitting System*, to analyze the security of the DLP with LHWP exponents.
We apply a parameterized splitting system to attack the GPS identification scheme modified by Coron, Lefranc and Poupard in CHES'05 and obtain an algorithm of $2^{61.6}$ time complexity which was expected to be 2^{78}. Also a parameterized splitting system can be used to solve the DLP with a LHWP exponent proposed by Hoffstein and Silverman in $2^{54.51}$ time complexity, that is smaller than 2^{59} in the recent Cheon-Kim attack.

Key words: Discrete Logarithm Problem with Low Hamming Weight Product (LHWP) Exponents, Parameterized Splitting Systems

1 Introduction

It is important to compute exponentiations efficiently in cryptosystems based on the DLP. One approach to achieve this is to choose an exponent of low Hamming weight. For example, the GPS identification scheme proposed by Girault [4, 5, 7] uses as a secret key a product of two integers having low Hamming weight [4, 5, 7]. Hoffstein and Silverman suggested a use of exponent $x = x_1x_2x_3$, where each integer x_i has very low Hamming weight [9]. But a use of low Hamming weight exponents may weaken the security.

The Heiman-Odlyzko algorithm [8] and the Coppersmith's splitting system [3, 10, 16] have been used to analyze the DLP with low Hamming weight exponents. The complexity of solving the DLP with the Coppersmith's splitting system is about the square root of the size of the key space when the exponent is a single integer. It can be regarded to be almost optimal since the DLP has the square root complexity in the generic model [14].

In [9], Hoffstein and Silverman proposed an attack against low Hamming weight product (LHWP) exponents. In [4], Coron, Lefranc and Poupard combined the above attack with the Coppersmith's splitting system and described

an algorithm that can be applied when the order of a group is unknown. But the complexity of the attack is far from the square root of the size of the key space.

Our results: In this paper, we generalize the Coppersmith's splitting system into a parameterized splitting system and propose its construction. It can be used to show that given a bit string of length n, weight t and a positive integer $t_1 < t$, there exists a part of the string of length n_1 and weight t_1 where $\frac{n_1}{t_1} \approx \frac{n}{t}$.

We apply a parameterized splitting system to the private key of the GPS identification scheme [4, 7] and the Hoffstein and Silverman's exponent [9] (originally designated for 2^{80} bit security). In [4], Coron, Lefranc and Poupard proposed an attack with 2^{52} complexity to recover the private key of the GPS identification scheme from CHES'04 and suggested a new private key which is claimed to have the security level of 2^{78}. But our parameterized splitting system reduces them to $2^{47.7}$ and $2^{65.5}$, respectively, and its randomized version reduces them to $2^{43.5}$ and $2^{61.6}$, respectively. In [1], Cheon and Kim introduced the notion of rotation-free elements and proposed an attack of $2^{55.9}$ complexity to the Hoffstein and Silverman's exponent. By combining the parameterized splitting system and the concept of rotation-freeness, we reduce it further to $2^{54.51}$.

Organization of the paper: In Section 2, we briefly introduce the Heiman-Odlyzko algorithm, the Coppersmith's splitting system and the rotation-free elements. In Section 3, we propose a parameterized splitting system and its application to the DLP of LHWP exponents. In Section 4, we analyze the complexity of the GPS identification scheme and the DLP with the Hoffstein and Silverman's exponent. Finally, we conclude in Section 5.

2 Preliminaries

Let g be a generator of a group G and x is an integer. From now on, *ord* g and $wt(x)$ denote the order of g and the Hamming weight of x, respectively.

Shanks' Baby-Step Giant-Step [13] and Pollard's Rho algorithm [11] are representative algorithms for the DLP. Algorithms for the DLP with low Hamming weight exponents are variants of Shanks' Baby-Step Giant-Step. In this section, we introduce the Heiman-Odlyzko algorithm, the Coppersmith's splitting system and the rotation-free elements. In this section, we assume *ord* g is known.

2.1 The Heiman-Odlyzko Algorithm

The Heiman-Odlyzko algorithm [8] was introduced by Heiman and Odlyzko independently. (In [8], Heiman remarked this algorithm was independently noticed by Odlyzko.) In this section, we sketch the Heiman-Odlyzko algorithm.

We use the notations from [16]. We regard the binary representation of

$$x = \sum_{i=0}^{n-1} x_i 2^i$$

as the vector

$$x = (x_0, \ldots, x_{n-1}).$$

Then this set of vectors corresponds to

$$\{i : x_i = 1\} \subset \mathbb{Z}_n.$$

The following two mappings, which are inverse to each other, express the above correspondence.

$$set: \ \{0, 1, \ldots, 2^n - 1\} \to 2^{\mathbb{Z}_n}, \ set(x = (x_0, \ldots, x_{n-1})) = \{i : x_i = 1\}$$
$$val: \ 2^{\mathbb{Z}_n} \to \{0, 1, \ldots, 2^n - 1\}, \ val(Y) = \sum_{i \in Y} 2^i$$

Consider the following equation

$$y = g^x = g^{x_1 + x_2},$$

where $t = wt(x) = wt(x_1) + wt(x_2)$, $wt(x_1) = t_s$ and $set(x_1) \cap set(x_2) = \emptyset$.

From the above equation, we get

$$yg^{-x_1} = g^{x_2}. \tag{1}$$

Now we compute yg^{-x_1} for all $x_1 \in \mathbb{Z}_n$ such that $wt(x_1) = t_s$ and build a lookup table that contains all the pairs (yg^{-x_1}, x_1) and support an efficient search on the first component. Then we compute g^{x_2} for each x_2 such that $wt(x_2) = t - t_s$ and look up the table until a collision is found.

Neglecting logarithmic factors, the time complexity of the Heiman-Odlyzko Algorithm is $O\left(\binom{n}{t_s} + \binom{n}{t-t_s}\right)$. Since we need store only either the left or the right hand side, the space complexity of the Heiman-Odlyzko Algorithm is $O\left(\min\{\binom{n}{t_s}, \binom{n}{t-t_s}\}\right)$.

2.2 The Coppersmith's Splitting System

The Coppersmith's splitting system was introduced in [10], based on the idea from [2]. Later, Stinson gave a good description of it in [16]. We follow this description.

Definition 1. *(The Splitting System)*
Suppose n and t are even integers, $0 < t < n$.[1] *A (n, t)-splitting system is a pair $(X, \mathcal{B})$ that satisfies the following properties.*
1. $|X| = n$ and $\mathcal{B}$ is a set of $\frac{n}{2}$-subsets of X called blocks.
2. For every $Y \subseteq X$ such that $|Y| = t$, there exists a block $B \in \mathcal{B}$ such that $|Y \cap B| = \frac{t}{2}$.

[1] Stinson constructed the splitting system even for odd n and t in [16].

Remark. An (n,t)-splitting system is denoted by an $(N;n,t)$-splitting system if it has N blocks.

The existence of a splitting system follows from this construction: Suppose $X = \mathbb{Z}_n = \{0,1,\ldots,n-1\},\ B_i = \{i+j \bmod n : 0 \le j \le \frac{n}{2}-1\},\ \mathcal{B} = \{B_i : 0 \le i \le \frac{n}{2}-1\}$. Then, $(X,\mathcal{B})$ is an $(\frac{n}{2};n,t)$-splitting system.

The Coppersmith's splitting system enables us to restrict to $\mathcal{B}$ the search space of x_1 and x_2 in Equation (1). Hence This algorithm requires $N\binom{\frac{n}{2}}{\frac{t}{2}}$ time complexity and $\binom{\frac{n}{2}}{\frac{t}{2}}$ space complexity.

A Randomized Algorithm The randomized version of the above algorithm is summarized in [16], which is also due to [3]. The time complexity of the randomized version is $O\left(\sqrt{t}\binom{\frac{n}{2}}{\frac{t}{2}}\right)$ and the space complexity of the randomized version is $O\left(\binom{\frac{n}{2}}{\frac{t}{2}}\right)$.

2.3 Rotation-Free Elements

In [1], Cheon and Kim defined an equivalent relation $\sim$ on $\mathbb{Z}_{2^n-1}$ as follows:

$a \sim b$ if and only if there exists a non-negative integer i such that $a = 2^i b$.

The idea of Cheon and Kim's attack on LHWPs is to reduce the key search space by considering only one element from each equivalent class.

Since there is no known algorithm to generate such representatives efficiently, they suggested a use of the set of *rotation-free elements* which contains at least one representative for each equivalent class. The set is only little bit larger than the number of equivalent classes and easily generated.

The definition of rotation-free elements is as follows:

Definition 2. *(Rotation-Free Elements [1])*
An element $z \in \mathbb{Z}_{2^n-1}$ is called a rotation-free element if there is a k-tuple $(a_1,a_2,\ldots,a_k)$ for a positive integer k satisfying

1. $a_i \ge a_1$ *for* $1 \le i \le k$.
2. $\sum_{i=1}^{k} a_i = n$.
3. $z = 2^{n-1} + 2^{n-1-a_1} + \cdots + 2^{n-1-(a_1+a_2+\cdots+a_{k-1})}$.

Let n,k be positive integers with $k < n$ and $RF(n,k)$ be the number of rotation-free elements of weight k in $\mathbb{Z}_{2^n-1}$. Then $RF(n,k)$ is given in [1] by

$$RF(n,k) = \sum_{i=0}^{\lfloor \frac{n}{k} \rfloor - 1} \binom{n-2-ki}{k-2}.$$

3 Parameterized Splitting Systems

In this section, we construct a *Parameterized Splitting System*, that is a generalization of the Coppersmith's splitting system. In the Coppersmith's splitting system, given $Y \subset \mathbb{Z}_n$, the size of a block B such that $|Y \cap B| = \frac{t}{2}$ is fixed to $\frac{n}{2}$. We show that the size of a block B can be flexible so that $|Y \cap B| = t_s$ and $|B| = \lfloor \frac{t_s n}{t} \rfloor$ for any $0 \leq t_s \leq t$. This flexibility yields an efficient algorithm for the DLP with LHWP exponents.

3.1 Parameterized Splitting Systems

We start with the definition of parameterized splitting systems.

Definition 3. *(Parameterized Splitting Systems)*
Suppose n and t are integers such that $0 < t < n$. For any t_s such that $0 \leq t_s \leq t$, a $(N; n, t, t_s)$-parameterized splitting system is a pair $(X, \mathcal{B})$ that satisfies the following properties.
1. $|X| = n$ and $\mathcal{B} = \{B \subset X : |B| = \lfloor \frac{t_s n}{t} \rfloor\}$.
2. $|\mathcal{B}| = N$.
3. For every $Y \subseteq X$ such that $|Y| = t$, there exists a block $B \in \mathcal{B}$ such that $|Y \cap B| = t_s$.

Remark. We may assume $0 < t < \frac{n}{2}$, $1 \leq t_s \leq \frac{t}{2}$.

The following Lemma 1 constructs an efficient parameterized splitting system.

Lemma 1. *$X = \{0, 1, \ldots, n-1\}$, $Y = \{y_1, y_2, \ldots, y_t\} \subset X$ such that $|Y| = t$. Suppose t_s is an integer such that $0 \leq t_s \leq t$. Let $B_i = \{i \bmod n, i + 1 \bmod n, \ldots, i + \lfloor \frac{t_s n}{t} \rfloor - 1 \bmod n\}$, $i = 0, 1, \ldots, n-1$. Then, there exists i such that $|Y \cap B_i| = t_s$.*

Proof. For each $y \in Y$, let $\nu(y) = \{i : y \in B_i,\ i = 0, 1, \ldots, n-1\}$. Then, $|\nu(y)| = \lfloor \frac{t_s n}{t} \rfloor$.

Let M be $\frac{1}{n} \sum_{i=0}^{n-1} |Y \cap B_i|$. Since $Y \cap B_i = \bigcup_{y \in Y} (\{y\} \cap B_i)$ and if $y_i \neq y_j$, then $(y_i \cap B_i) \cap (y_j \cap B_i) = \emptyset$,

$$M = \frac{1}{n} \sum_{I=0}^{n-1} |Y \cap B_i| = \frac{1}{n} \sum_{i=0}^{n-1} |\bigcup_{y \in Y} (\{y\} \cap B_i)| = \frac{1}{n} \sum_{i=0}^{n-1} \sum_{y \in Y} |\{y\} \cap B_i|$$
$$= \frac{1}{n} \sum_{y \in Y} \sum_{i=0}^{n-1} |\{y\} \cap B_i| = \frac{1}{n} \sum_{y \in Y} |\nu(y)| = \frac{t}{n} \left\lfloor \frac{t_s n}{t} \right\rfloor.$$

From $\frac{t_s n}{t} - 1 < \lfloor \frac{t_s n}{t} \rfloor \leq \frac{t_s n}{t}$,

$$t_s - 1 < t_s - \frac{t}{n} = \frac{t}{n} \cdot (\frac{t_s n}{t} - 1) < \frac{t}{n} \left\lfloor \frac{t_s n}{t} \right\rfloor = M \leq \frac{t}{n} \cdot \frac{t_s n}{t} = t_s. \qquad (2)$$

Suppose there doesn't exist B_i such that $|Y \cap B_i| = t_s$. If $|Y \cap B_i| < t_s$ for all i, then $M \leq t_s - 1$, which contradicts with Equation (2). If $|Y \cap B_i| > t_s$ for all i, then $t_s + 1 \leq M$, which contradicts with Equation (2).

By the above discussions, there exists B_i and B_j such that $|Y \cap B_i| \leq t_s$ and $|Y \cap B_j| \geq t_s$. However, from the fact $|Y \cap B_i| - |Y \cap B_{i+1}| \in \{-1,\ 0,\ 1\}$, $|Y \cap B_k|$ should be t_s for some $k \in \{i \bmod n, i+1 \bmod n, \ldots, j-1 \bmod n, j \bmod n\}$, which contradicts with the assumption.

Therefore, there exists B_i such that $|Y \cap B_i| = t_s$. □

Theorem 1. *Let* $X = \{0, 1, \ldots, n-1\}$, $B_i = \{i \bmod n, i+1 \bmod n, \ldots, i + \lfloor \frac{t_s n}{t} \rfloor - 1 \bmod n\}$, $\mathcal{B} = \{B_i : 0 \leq i \leq n-1\}$. *Then,* $(X, \mathcal{B})$ *is a* $(n; n, t, t_s)$*-parameterized splitting system.*

A Randomized Version For given Y and t_s, Theorem 1 implies that if we try at most n blocks, we can find some block B such that $|Y \cap B| = t_s$. In a randomized version, we randomly choose $B \subset \mathbb{Z}_n$ such that $|B| = \lfloor \frac{t_s n}{t} \rfloor$ and check whether $|Y \cap B| = t_s$. Then the probability of success is

$$p = \frac{\binom{t}{t_s}\binom{n-t}{\lfloor \frac{t_s n}{t} \rfloor - t_s}}{\binom{n}{\lfloor \frac{t_s n}{t} \rfloor}}.$$

Lemma 3 shows that the expected number of trials to find a good block B such that $|Y \cap B| = t_s$ is $O(\sqrt{t})$. We require Lemma 2 from [16] to get Lemma 3.

Lemma 2. *Suppose that* n *and* λn *are positive integers, where* $0 < \lambda < 1$. *Define*

$$H(\lambda) = \lambda \log_2 \lambda - (1-\lambda)\log_2(1-\lambda).$$

Then

$$\frac{2^{nH(\lambda)}}{\sqrt{8n\lambda(1-\lambda)}} \leq \binom{n}{\lambda n} \leq \frac{2^{nH(\lambda)}}{\sqrt{2\pi n\lambda(1-\lambda)}}.$$

Lemma 3. $p > \sqrt{\frac{\pi}{2}} \cdot \sqrt{\left(\frac{t_s}{t} - \frac{1}{n}\right)\left(1 - \frac{t_s}{t}\right)} \cdot t^{-1/2} \geq \frac{\sqrt{\pi}}{4} t^{-1/2}$.

Proof.

$$p = \binom{t}{t_s} \frac{\binom{n-t}{\lfloor \frac{t_s n}{t} \rfloor - t_s}}{\binom{n}{\lfloor \frac{t_s n}{t} \rfloor}} = \binom{t}{\lambda_1 t} \frac{\binom{n-t}{\lambda_2(n-t)}}{\binom{n}{\lambda n}},$$

where $\lambda_1 = \frac{t_s}{t}$, $\lambda_2 = \frac{\lfloor \frac{t_s n}{t} \rfloor - t_s}{n-t}$ and $\lambda = \frac{\lfloor \frac{t_s n}{t} \rfloor}{n}$.

From Lemma 2,

$$\begin{aligned} p &\geq \frac{2^{tH(\lambda_1)}}{\sqrt{8t\lambda_1(1-\lambda_1)}} \cdot \frac{2^{(n-t)H(\lambda_2)}}{\sqrt{8(n-t)\lambda_2(1-\lambda_2)}} \cdot \frac{\sqrt{2\pi n\lambda(1-\lambda)}}{2^{nH(\lambda)}} \\ &= \frac{2^{tH(\lambda_1)+(n-t)H(\lambda_2)}}{2^{nH(\lambda)}} \cdot \frac{\sqrt{2\pi n\lambda(1-\lambda)}}{8\sqrt{t(n-t)\lambda_1(1-\lambda_1)\lambda_2(1-\lambda_2)}}. \end{aligned}$$

Since $H(\lambda)$ is convex,

$$tH(\lambda_1) + (n-t)H(\lambda_2) \geq nH(\lambda),$$

hence,

$$p \geq \frac{\sqrt{2\pi n\lambda(1-\lambda)}}{8\sqrt{t(n-t)\lambda_1(1-\lambda_1)\lambda_2(1-\lambda_2)}}.$$

Since $0 < \lambda_i < 1$,

$$\frac{1}{\sqrt{\lambda_i(1-\lambda_i)}} \geq 2$$

for $i = 1, 2$, hence,

$$p \geq \sqrt{\frac{\pi}{2}} \cdot \sqrt{\lambda(1-\lambda)} \cdot t^{-1/2}.$$

We may assume $1 \leq t_s \leq \frac{t}{2}$ and $2 \leq t \leq \frac{n}{2}$. From $\lambda = \lfloor \frac{t_s n}{t} \rfloor / n$, we have

$$\lambda(1-\lambda) > (\frac{t_s}{t} - \frac{1}{n})(1 - \frac{t_s}{t}) \geq \frac{1}{8}.$$

□

3.2 The DLP with LHWP Exponents when the Order of g is Known

Before detailing how parameterized splitting systems can be used, we review some known methods.

For an integer x, we denote by $|x|$ the bit-length of x. Let $X_1 = \{x_1 : |x_1| = n_1,\ wt(x_1) = t_1\}$ and $X_2 = \{x_2 : |x_2| = n_2,\ wt(x_2) = t_2\}$. Consider $x{=}x_1x_2$, where $x_1 \in X_1$ and $x_2 \in X_2$.

As in [4, 9], from the following equation

$$y = (g^{x_1})^{x_2} = h^{x_2},$$

x can be computed by repeating an algorithm for the DLP by $|X_1|$. So, the time complexity and the space complexity of the Heiman-Odlyzko algorithm are

$$O\left(|X_1|\left(\binom{n_2}{t_s} + \binom{n_2}{t-t_s}\right)\right) \text{ and } O\left(\min\left\{\binom{n_2}{t_s}, \binom{n_2}{t-t_s}\right\}\right),$$

respectively. To minimize the time complexity, t_s should be $\lceil \frac{t_2}{2} \rceil$ or $\lfloor \frac{t_2}{2} \rfloor$. The time complexity and the space complexity of the parameterized splitting system are

$$O\left(|X_1| \cdot n_2 \binom{\frac{n_2}{2}}{\frac{t_2}{2}}\right) \text{ and } O\left(\binom{\frac{n_2}{2}}{\frac{t_2}{2}}\right),$$

respectively.

Another attack, which is also followed from [4, 9], takes the trade-off between time and space. $y = g^{x_1 x_2}$ can be converted into

$$y^{{x_1}^{-1}} g^{-x_3} = g^{x_4},$$

where $x_2 = x_3 + x_4$ and $set(x_3) \cap set(x_4) = \emptyset$. Note that x_1^{-1} denotes the multiplicative inverse of x_1 modulo the order of g.

Put $wt(x_3) = t_s$. From the above equation, we find x_1 and x_2 by computing both sides and comparing them.

Therefore the time complexity and the space complexity of the Heiman-Odlyzko algorithm are

$$O\left(|X_1|\binom{n_2}{t_s} + \binom{n_2}{t - t_s}\right) \text{ and } O\left(\min\left\{|X_1|\binom{n_2}{t_s}, \binom{n_2}{t - t_s}\right\}\right),$$

respectively. t_s is an integer such that $0 \leq t_s \leq \lceil \frac{t_2}{2} \rceil$. Comparing to the first application, the time complexity is lower.

The time complexity and the space complexity of the splitting system are

$$O\left(|X_1| \cdot \frac{n_2}{2}\binom{\frac{n_2}{2}}{\frac{t_2}{2}} + \frac{n_2}{2}\binom{\frac{n_2}{2}}{\frac{t_2}{2}}\right) = O\left(|X_1| \cdot \frac{n_2}{2}\binom{\frac{n_2}{2}}{\frac{t_2}{2}}\right) \text{ and } O\left(\frac{n_2}{2}\binom{\frac{n_2}{2}}{\frac{t_2}{2}}\right),$$

respectively. Comparing to the first application, the efficiency of the time complexity is hardly improved.

In the case of the DLP with a single integer exponent of low Hamming weight, the splitting system appears to be more efficient than the Heiman-Odlyzko algorithm since one of the factors of the time complexity, n_2, is reduced to $\frac{n_2}{2}$ in the splitting system. But the splitting system fixes $t_s = \frac{t_2}{2}$ while the Heiman-Odlyzko algorithm is able to choose t_s arbitrary. This difference yields the Heiman-Odlyzko algorithm carries out trade-off efficiently while the splitting system does not.

Now we propose a new algorithm using parameterized splitting systems, which takes the advantages from both of previous algorithms. From Section 3.1, for $t_s \in [0, \lceil \frac{t_2}{2} \rceil]$, there exists a $(n_2; n_2, t_2, t_s)$-parameterized splitting system $(\mathbb{Z}_{n_2}, \mathcal{B})$. So, there is a block $B_i \in \mathcal{B}$ such that $|set(x_2) \cap B_i| = t_s$. Let

$$set(x_3) = set(x_2) \cap B_i \text{ and } set(x_4) = set(x_2) \cap (\mathbb{Z}_{n_2} - B_i).$$

Then, we get the following equation

$$y^{{x_1}^{-1}} g^{-val(set(x_2) \cap B_I)} = g^{val(set(x_2) \cap (\mathbb{Z}_{n_2} - B_i))}.$$

From the above equation, we get Algorithm 1. The first part of Algorithm 1 is to compute and store all the values of the left-hand side. The second part of Algorithm 1 is to compute each value of the right-hand side and check if it is in the list from the first part.

Now we present Algorithm 1 and its randomized version.

Algorithm 1
Finding discrete logarithm when the order of g is known (deterministic)

Input: $g,\ y \in G,\ X_1,\ (n_2; n_2, t_2, t_s)$-parameterized splitting system $(\mathbb{Z}_{n_2}, \mathcal{B})$
Output: $\log_g y$

for all $x_1 \in X_1$ **do**
  for all B_i **do**
    for all $Y_{1,i} \subset B_i$ such that $|Y_{1,i}| = t_s$ **do**
      Compute ${y^{x_1}}^{-1} g^{-val(Y_{1,i})}$
      Add $(x_1, val(Y_{1,i}), {y^{x_1}}^{-1} g^{-val(Y_{1,i})})$ to the list L
      Sort L by third coordinate
    end for
  end for
end for
for all $\mathbb{Z}_{n_2} - B_i$ **do**
  for all $Y_{2,i} \subset \mathbb{Z}_{n_2} - B_I$ such that $|Y_{2,i}| = t_2 - t_s$ **do**
    Compute $g^{val(Y_{2,i})}$
    if $g^{val(Y_{2,i})}$ is the third coordinate of some entry in the list L **then**
      return $x_1(val(Y_{1,i}) + val(Y_{2,i}))$
    end if
  end for
end for

Algorithm 2
Finding discrete logarithm when the order of g is known (randomized)

Input: $g,\ y \in G,\ X_1,\ t_s$
Output: $\log_g y$

loop
  Choose randomly $B \subset \mathbb{Z}_{n_2}$ such that $|B| = \lfloor \frac{t_s n_2}{t_2} \rfloor$
  for all $x_1 \in X_1$ **do**
    for all $Y_1 \subset B$ such that $|Y_1| = t_s$ **do**
      Compute ${y^{x_1}}^{-1} g^{-val(Y_1)}$
      Add $(x_1, val(Y_1), {y^{x_1}}^{-1} g^{-val(Y_1)})$ to the list L
      Sort L by third coordinate
    end for
  end for
  for all $Y_2 \subset \mathbb{Z}_{n_2} - B$ such that $|Y_2| = t_2 - t_s$ **do**
    Compute $g^{val(Y_2)}$
    if $g^{val(Y_2)}$ is the third coordinate of some entry in the list L **then**
      return $x_1(val(Y_1) + val(Y_2))$
    end if
  end for
end loop

Analysis: Algorithm 1 needs $|X_1| \cdot n_2 \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s}$ exponentiations in the first part and $n_2 \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}$ exponentiations in the second part. In Algorithm 1, we can store $(val(Y_{2,i}),\ g^{val(Y_{2,i})})$'s instead of $(x_1,\ val(Y_{1,i}),\ {y^{x_1}}^{-1} g^{-val(Y_{1,i})})$'s. In this case, we compute ${y^{x_1}}^{-1} g^{-val(Y_{1,i})}$ and find a collision. So, we store one of two sets which has smaller cardinality. Thus, the time complexity and the space complexity (neglecting logarithmic factors) are

$$O\left(|X_1| \cdot n_2 \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s} + n_2 \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right) \text{ and}$$
$$O\left(\min\left\{|X_1| \cdot n_2 \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s},\ n_2 \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right\}\right),$$

respectively.

Lemma 3 implies that in about $\frac{4}{\sqrt{\pi}} {t_2}^{1/2}$ iterations Algorithm 2 outputs $\log_g y$. And we only make L for each B. Thus, if we count the number of group exponentiations, the time complexity and the space complexity are

$$O\left(|X_1| \cdot \sqrt{t_2} \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s} + \sqrt{t_2} \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right) \text{ and}$$
$$O\left(\min\left\{|X_1| \cdot \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s},\ \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right\}\right),$$

respectively.

3.3 The DLP with LHWP Exponents when the Order of g is Unknown

Recall the following equation in Section 3.2,

$$ {y^{x_1}}^{-1} g^{-x_3} = g^{x_4}, \tag{3}$$

If *ord* g is unknown, x_1^{-1} is not easy to compute from x_1 and so Equation (3) cannot be checked directly.

However, we can use Algorithm 1 or 2 from following trick from [4] and, earlier, proposed by Shoup [15]. Let

$$\chi = \prod_{x \in X_1} x \text{ and } \hat{g} = g^{\chi}.$$

From

$$({y^{x_1}}^{-1} g^{-x_3})^{\chi} = (g^{x_4})^{\chi},$$

we get

$$ {y^{\prod_{x \in X_1 - \{x_1\}} x}} \cdot \hat{g}^{-x_3} = \hat{g}^{x_4}, \tag{4}$$

where $x_2 = x_3 + x_4$ and $set(x_3) \cap set(x_4) = \emptyset$.

To solving the DLP, we should perform the precomputation of $y^{\prod_{x \in X_1 - \{x_1\}} x}$, $\hat{g}$ and $\hat{g}^{-1}$ and store them.

$\{y^{\prod_{x \in X_1 - \{x_1\}} x} : x_i \in X_1\}$ can be computed by the algorithm proposed by Coron, Lefranc and Poupard in [4]. According to the algorithm, $|X_1| \cdot \log_2 |X_1|$ group exponentiations are necessary.

Therefore if we are able to learn $\hat{g}^{-1}$, we have Algorithm 3 and Algorithm 4.

Algorithm 3
Finding discrete logarithm when the order of g is unknown (deterministic)

Input: $g,\ y \in G,\ X_1,\ (n_2; n_2, t_2, t_s)$-parameterized splitting system $(\mathbb{Z}_{n_2}, \mathcal{B})$
Output: $\log_g y$

1: Compute $y^{\prod_{x \in X_1 - \{x_1\}} x}$, $\hat{g}$ and $\hat{g}^{-1}$ and store them
2: Substituting $\hat{g}$ for $g,\ \hat{g}^{-1}$ for g^{-1} and $\{y^{\prod_{x \in X_1 - \{x_1\}} x} : x_i \in X_1\}$ for X_1, carry out Algorithm 1

Algorithm 4
Finding discrete logarithm when the order of g is unknown (randomized)

Input: $g,\ y \in G,\ X_1$
Output: $\log_g y$

1: Compute $y^{\prod_{x \in X_1 - \{x_1\}} x}$, $\hat{g}$ and $\hat{g}^{-1}$ and store them
2: Substituting $\hat{g}$ for $g,\ \hat{g}^{-1}$ for g^{-1} and $\{y^{\prod_{x \in X_1 - \{x_1\}} x} : x_i \in X_1\}$ for X_1, carry out Algorithm 2

Analysis: First, we analyze Algorithm 3. In Step 1, we perform $|X_1| \cdot \log_2 |X_1|$ group exponentiations and store the results. There is no change of the time complexity and space complexity in Step 2. Therefore, the time complexity is

$$O\left(|X_1| \cdot \log_2 |X_1| + |X_1| \cdot n_2 \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s} + n_2 \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right)$$

and the space complexity is

$$O\left(|X_1| \cdot \log_2 |X_1| + \min\left\{|X_1| \cdot n_2 \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s},\ n_2 \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right\}\right).$$

The best efficiency of the time complexity can be achieved when $|X_1|\binom{\lfloor \frac{t_s n_2}{t} \rfloor}{t_s} \approx \binom{n_2 - \lfloor \frac{t_s n_2}{t} \rfloor}{t_2 - t_s}$. At this t_s, $|X_1| \cdot \log_2 |X_1|$ is negligible.

The only difference with Algorithm 3 is Step 2. Therefore, the time complexity is

$$O\left(|X_1| \cdot \log_2 |X_1| + |X_1| \cdot \sqrt{t_2}\binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s} + \sqrt{t_2}\binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right)$$

and the space complexity is

$$O\left(|X_1| \cdot \log_2 |X_1| + \min\left\{|X_1| \cdot \binom{\lfloor \frac{t_s n_2}{t_2} \rfloor}{t_s}, \binom{n_2 - \lfloor \frac{t_s n_2}{t_2} \rfloor}{t_2 - t_s}\right\}\right).$$

Remark. We note that Algorithm 3 and 4 might output false answers. These errors come from the fact that the order of $\hat{g}$ of Equation (4) might be smaller than that of g. The worst case is that the order of g is a divisor of that of $\hat{g}$. In this case, Equation (4) is an identical equation.

4 Applications

In this section, we attack the private keys of the GPS identification scheme [5, 6, 12] and the exponent proposed by Hoffstein and Silverman [9].

4.1 Attacks on Private Keys of the GPS Identification Scheme

We briefly introduce the GPS identification scheme.

GPS Identification Scheme The GPS identification scheme, such as labelled by the NESSIE project, is an interactive protocol between a prover and a verifier which contains one or several rounds of three passes [7]. The GPS identification scheme is based on the DLP over ${\mathbb{Z}_N}^*$. Precisely, when g is an element of ${\mathbb{Z}_N}^*$ of maximal order m, the GPS identification scheme is based on the DLP over $G = \langle g \rangle$, where *ord* g is secret. When $y = g^{-x} \bmod N$, a private key of a prover is x and public keys are $(N,\ g,\ y)$. N is the product of two primes and the factorization of N should be difficult.

There are four security parameters as follows:

I. S is the binary size of x. Typically, S=160.
ii. k is the binary size of the challenges sent to the prover and determines the level of security of the scheme.
iii. R is the binary size of the exponents used in the commitment computation. It typically verifies $R = S + k + 80$.
iv. m is the number of rounds the scheme is iterated. Theoretically, m is polynomial in the size of the security parameter. But, in practice, m is often chosen equal to 1.

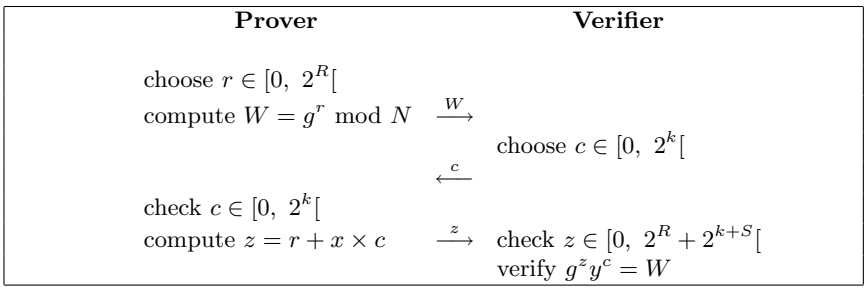

Fig 1. The GPS Identification Scheme

Private Keys of the GPS Identification Scheme For the efficiency of the protocol, Girault and Lefranc proposed a private key x as $x = x_1 x_2$ in [7], where x_1 is a 19-bit number with 5 random bits equal to 1 chosen among the 16 least significant ones, x_2 is a 142-bit number with 16 random bits equal to 1 chosen among the 138 least significant ones in CHES'04.

Later in CHES'05, to strengthen the security, Coron, Lefranc and Poupard suggest the modified x_1 and x_2 in [4], where x_1 is a 30-bit number with 12 nonzero bits and x_2 is a 130-bit number with 26 nonzero bits.

Attacks on Private Keys We put $|X_1| = \binom{16}{5},\ n_2 = 138,\ t_2 = 16$ for private keys from [7] and $|X_1| = \binom{30}{12},\ n_2 = 130,\ t_2 = 26$ for private keys from [4]. Since N is public we can easily compute $\hat{g}^{-1}$ of Algorithm 2, using the extended Euclidean algorithm. Before applying these private keys to Algorithm 3 and Algorithm 4, we note that when t_s is chosen to guarantee the most efficient time complexity, the cost of precomputation is negligible.

Table 1 compares the complexities of recovering private keys from [7] and Table 2 for [4]. The private key from [7] was broken in [4], which needs 2^{52} group exponentiations. But the parameterized splitting system and its randomized version reduce it further to $2^{47.7}$ and $2^{43.5}$, respectively.

Method	Exponentiations	Storage
[7]	2^{52}	2^{33}
Ours (Algorithm 3), $t_s = 7$	$2^{47.7}$	$2^{44.5}$
Ours (Algorithm 4), $t_s = 7$	$2^{43.5}$	2^{41}

Table 1. Private Keys from [7]

Table 2 shows that the parameterized splitting system and its randomized version reduce the complexity of the DLP with the private key proposed in [4] from 2^{78} to $2^{65.5}$ and $2^{62.1}$, respectively.

Method	Exponentiations	Storage
[4]	2^{78}	$2^{43.9}$
Ours (Algorithm 3), $t_s = 9$	$2^{65.5}$	$2^{63.1}$
Ours (Algorithm 4), $t_s = 9$	$2^{61.6}$	$2^{59.2}$

Table 2. Private Keys from [4]

4.2 Attacks on the Hoffstein and Silverman's Exponent

The Hoffstein and Silverman's Exponent Hoffstein and Silverman proposed a use of exponent $x = x_1x_2x_3 \in \mathbb{Z}_{2^{1000}-1}$, where x_1, x_2 and x_3 are integers of $wt(x_1) = 6$, $wt(x_2) = 7$ and $wt(x_3) = 7$ or $wt(x_1) = 2$, $wt(x_2) = 2$ and $wt(x_3) = 11$ [9]. In the case of $wt(x_1) = 6$, $wt(x_2) = 7$ and $wt(x_3) = 7$, all values of the Hamming weight are similar, hence, splitting of one's Hamming weight doesn't give advantages. So we focus on the case of $wt(x_1) = 2$, $wt(x_2) = 2$ and $wt(x_3) = 11$.

Let $y = g^x$ for $x = x_1x_2x_3$ where x_i's are of weight (2,2,11). Following the trick in [1], we rewrite x as $x = 2^k\bar{x}_1\bar{x}_2x_3$ where $0 \le k < n$ and each of $\bar{x}_i$ are rotation-free elements in the same equivalent class with x_i for each i. We further split x_3 by $x_3 = x_3' + x_3''$ where x_3' and x_3'' have weight 3 and 8, respectively. Then we can find x by checking the following equations:

$$y^{2^{-k}\bar{x}_1^{-1}\bar{x}_2-1}g^{-x_3'} = g^{x_3''}.$$

In [1], Cheon and Kim modify k so that x_3'' becomes rotation-free. Then the complexity for $n = 1000$ is

$$n \cdot RF(n,2)^2\binom{n-1}{3} + RF(n,8) \approx 2^{55.2} + 2^{54.5} \approx 2^{55.9}.$$

On the other hand, if we combine the existence of a parameterized splitting system and the notion of the rotation-free, we get a little bit smaller complexity. When we split x_3, we apply the Theorem 1 to find a block B such that $|B| = \lfloor \frac{3n}{11} \rfloor$ and $|set(x_3) \cap B| = 3$. We write $set(x_3) \cap (\mathbb{Z}_n - B) = \{s_0, s_1, \ldots, s_7\}$ and let l_i be the number of elements of $\mathbb{Z}_n$ in $[s_i, s_{i+1}]$ for $i = 0, 1, \ldots, 7$, where we set $s_8 = s_1$ and $[s_7, s_1] = \{s_7, \ldots, n-1, 0, \ldots, s_1\}$. Suppose l_j is the maximum of l_i's. Then, l_j should be larger than $\lfloor \frac{3n}{11} \rfloor$. We shift x_3 so that s_j is placed at 0.

From the above discussions, there exists an integer k' such that $2^{k'}x_3 = x_3' + x_3''$, where x_3' and x_3'' satisfy

1. x_3' is a string of length n and weight 3. If we write $set(x_3') = \{a_0, a_1, a_2\}$ for $0 < a_0 < a_1 < a_2 \le n-1$, then $a_2 - a_0 + 1 \le \lfloor \frac{3n}{11} \rfloor$.
2. x_3'' is a string of length n and weight 8. If we write $set(x_3'') = \{b_0, b_1, \ldots, b_7\}$ for $0 = b_0 < b_1 < \cdots < b_7 \le n-1$, then $b_i - b_{i-1} \le b_1$ and $\lfloor \frac{3n}{11} \rfloor \le b_1$.

To enumerate the number N_1 of x_3', we first fix $a_0 \in [1, n-3]$ and then choose distinct $a_1, a_2 \in [a_0+1, \min\{a_0 - 1 + \lfloor \frac{3n}{11} \rfloor, n-1\}]$. Hence

$$N_1 = \sum_{a_0=1}^{\lceil \frac{8n}{11} \rceil} \binom{\lfloor \frac{3n}{11} \rfloor - 1}{2} + \sum_{a_0=\lceil \frac{8n}{11} \rceil + 1}^{n-3} \binom{n-1-a_0}{2}.$$

To enumerate the number N_2 of x_3'', we let $l_0 = b_1$, $l_i = b_{i+1} - b_i$ for $i = 1, \ldots, 6$ and $l_7 = n - 1 - b_7$. Then, N_2 is the number of 8-tuple $(l_0, \ldots, l_7)$ satisfying

1. $\sum_{i=0}^{7} l_i = n - 1$.
2. $\lfloor \frac{3n}{11} \rfloor \le l_0 \le n - 7$.
3. $1 \le l_i \le l_0$ for $i = 1, \ldots, 6$ and $0 \le l_7 \le l_0$.

First, we enumerate the number of solutions satisfying the above conditions when $l_7 \neq 0$. Consider the following equation.

$$\sum_{i=1}^{7} l_i = n - 1 - l_0. \tag{5}$$

This is the problem that how many solutions of positive integers the linear Diophantine equation (5) has when $1 \le l_i \le 0$ for $i = 1, \cdots, 6$.

Given l_0, Let $A(l_0)$ be the set of solutions of Equation (5), $A_i(l_0)$ be the set of solutions when $l_i > l_0$ and $A_{i,j}(l_0)$ be the set of solutions when $l_i > l_0$ and $l_j > l_0$. Note that when $\lfloor \frac{3n}{11} \rfloor \le l_0 \le \lfloor \frac{n-2}{3} \rfloor$, only up to two values of l_i, $i = 1, \cdots, 7$ can be larger than l_0, because otherwise, the sum of the others should be less than 0. Similarly, when $\lfloor \frac{n-2}{3} \rfloor + 1 \le l_0 \le \lfloor \frac{n-2}{2} \rfloor$, only one value can be larger than l_0 and when $\lfloor \frac{n-2}{2} \rfloor + 1 \le l_0 \le n - 7$, any value cannot be larger than l_0. Thus for given l_0, the number of solutions in the case of $l_7 \neq 0$ is

$$\begin{aligned} N_{2,1}(l_0)' &= |A(l_0)| - |\bigcup_{i=0}^{7} A_i(l_0)^c| = |A(l_0)| - \{\sum_{i=0}^{7} |A_i(l_0)| - \sum_{i \neq j} |A_{i,j}(l_0)|\} \\ &= \binom{n-2-l_0}{6} - \left\{ 7\binom{n-2-2l_0}{6} - \binom{7}{2}\binom{n-2-3l_0}{6} \right\} \end{aligned}$$

when $\lfloor \frac{3n}{11} \rfloor \le l_0 \le \lfloor \frac{n-2}{3} \rfloor$. When $\lfloor \frac{n-2}{3} \rfloor + 1 \le l_0 \le \lfloor \frac{n-2}{2} \rfloor$,

$$\begin{aligned} N_{2,2}(l_0)' &= |A(l_0)| - |\bigcup_{i=0}^{7} A_i(l_0)^c| = |A(l_0)| - \sum_{i=0}^{7} |A_i(l_0)| \\ &= \binom{n-2-l_0}{6} - 7\binom{n-2-2l_0}{6}. \end{aligned}$$

When $\lfloor \frac{n-2}{2} \rfloor + 1 \le l_0 \le n - 7$,

$$N_{2,3}(l_0)' = |A(l_0)| = \binom{n-2-l_0}{6}.$$

When $l_7 = 0$, the number of solutions $N_{2,i}(l_0)''$, $i = 1, 2, 3$, can be computed in a similar way, *i.e.,* 6 in each binomial is replaced to 5.

Thus,

$$N_2 = \sum_{l_0=\lfloor \frac{3n}{11} \rfloor}^{\lfloor \frac{n-2}{3} \rfloor} (N_{2,1}(l_0)' + N_{2,1}(l_0)'') + \sum_{l_0=\lfloor \frac{n-2}{3} \rfloor+1}^{\lfloor \frac{n-2}{2} \rfloor} (N_{2,3}(l_0)' + N_{2,3}(l_0)'') + \sum_{\lfloor \frac{n-2}{2} \rfloor+1}^{n-7} (N_{2_3}(l_0)' + N_{2,3}(l_0)'').$$

Therefore, the total time complexity of the combined algorithm is

$$n \cdot RF(n,2)^2 N_1 + N_2 \approx 2^{52.75} + 2^{54.01} \approx 2^{54.51}.$$

And the space complexity of the combined algorithm is about $2^{52.75}$.

5 Conclusion

In this paper, we have proposed a *Parameterized Splitting System* and its randomized version. Since a parameterized splitting system takes the advatages from both of the splitting system and the Heiman-Odlyzko algorithm, it gives an efficient algorithm for the DLP with LHWP exponents.

Acknowledgements The authors would like to thank Martijn Stam and the anonymous referees for valuable comments. The first author also would like to thank Namsu Jho for helpful discussions. This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. R11-2007-035-01002-0).

References

1. J. Cheon and H. Kim, *Analysis of Low Hamming Weight Products,* To appear in Discrete Applied Mathematics.
2. D. Coppersmith and G. Seroussi, *On the Minimum Distance of Some Quadratic Residue Codes,* IEEE Trans. Inform. Theory 30 (1984), MR 86c:94025, pp 407–411.
3. D. Coppersmith, *Private communication to Scott Vanstone*, December 1997.
4. J. Coron, D. Lefranc and G. Poupard, *A New Baby-Step Giant-Step Algorithm and Some Application to Cryptanalysis,* Proc CHES 2005, LNCS 3656, Springer-Verlag, 2005, pp. 47–60.
5. M. Girault, *Self-Certified Public Keys,* Proc. Eurocrypt 1991, LNCS 547, Springer-Verlag, 1991, pp. 490–497.
6. M. Girault, G. Poupard and J. Stern, *Some Modes of Use of the GPS Idetification Scheme,* 3rd Nessie Conference, Springer-Verlag, November 2002.
7. M. Girault and D. Lefranc, *Public Key Authentication with One Single (on-line) Addition,* Proc. CHES 2004, LNCS 3156, Springer-Verlag, 2004, pp. 413–427.

8. R. Heiman, *A Note on Discrete Logarithms with Special Structure,* Proc. Eurocrypt 1992, LNCS 658, Springer-Verlag, 1993, pp 454–457.
9. J. Hoffstein and J. Silverman, *Random Small Hamming Weight Products with Application to Cryptography,* Discrete Appl. Math., Vol. 130, No.1, 2003, pp. 37–49.
10. A. Menezes, P. van Oorschot and S. Vanstone, *Handbook of Applied Cryptography,* CRC Press, 1997, pp 128.
11. J. Pollard, *Monte Carlo Methods for Index Computation (mod p),* Mathematics of Computation, Vol. 32, No. 143, 1978, pp 918–924.
12. G. Poupard and J. Stern, *Security Analysis of a Practical "On the Fly" Authentication and Signature Generation,* Proc. Eurocrypt 1998, LNCS 1403, Springer-Verlag, 1998, pp 422–436.
13. D. Shanks. *Class Number, a Theory of Factorization and Genera,* Proc. Symp. Pure Math., Vol. 20, 1971, pp 415–440.
14. V. Shoup, *Lower Bounds for discrete Logarithms and Related Problems,* Proc. Eurocrypt 1997, LNCS 1233, Springer-Verlag, 1997, pp 256–266.
15. V. Shoup, *Practical Threshold Signatures,* Proc. Eurocrypt 2000, LNCS 1807, Springer-Verlag, 2000, pp 207–220.
16. D. Stinson. *Some Baby-Step Giant-Step Algorithms for the Low Hamming Weight Discrete Logarithm Problem,* Mathematics of Computation, Vol. 71, No. 237, 2002, pp 379–391.