Exact Analysis of Montgomery Multiplication

Hisayoshi Sato[1], Daniel Schepers*[2], and Tsuyoshi Takagi[2]

[1] Hitachi, Ltd., Systems Development Laboratory,
292, Yoshida-cho, Totsuka-ku, Yokohama, 244-0817, Japan
hisato@sdl.hitachi.co.jp
[2] Technische Universität Darmstadt, Fachbereich Informatik,
Hochschulstr.10, D-64283 Darmstadt, Germany
{schepers, takagi}@informatik.tu-darmstadt.de

Abstract. The Montgomery multiplication is often used for an efficient implementations of public-key cryptosystems. This algorithm occasionally needs an extra subtraction in the final step, and the correlation of these subtractions can be considered as an invariant of the algorithm. Some side channel attacks on cryptosystems using Montgomery Multiplication has been proposed applying the correlation estimated heuristically. In this paper, we theoretically analyze the properties of the final subtraction in Montgomery multiplication. We investigate the distribution of the outputs of multiplications in the fixed length interval included between 0 and the underlying modulus. Integrating these distributions, we present some proofs with a reasonable assumption for the appearance ratio of the final subtraction, which have been heuristically estimated by previous papers. Moreover, we present a new invariant of the final subtraction: $x \cdot y$ with $y = 3x \bmod m$, where m is the underlying modulus. Finally we show a possible attack on elliptic curve cryptosystems using this invariant.

Keywords: timing attack, elliptic curve cryptosystem, Montgomery multiplication, randomization.

1 Introduction

The Montgomery Multiplication is widely utilized in implementations for public-key cryptosystems [9]. The Montgomery multiplication is an efficient algorithm for computing modular multiplication without the use of relatively expensive division with remainder, and it is suitable for the memory-constraint devices such as smart cards.

Since 1996 timing attacks gained more and more interest. After Kocher [6, 7] started with the first attacks on DSS and RSA numerous researchers worked on this topic. RSA and DES were probably the targets which have been attacked most. This kind of attack is especially attractive to smart cards. Dhem et al. proposed the first timing attack on RSA using Montgomery multiplication [4]. They focused on the final subtraction which appears in the Montgomery multiplication. They experimentally showed a timing attack by analyzing the distribution

* The second author is supported by SicAri Project (www.sicari.de) — German Federal Ministry of Education and Research.

of the appearance ratio correlated to the secret information. From their experiment the appearance ratio is about 17% on average.

After the timing attack, some theoretical analysis about the final subtraction have been investigated. Schindler heuristically showed a relationship between the appearance ratio and the underlying parameters [10]. He estimated the appearance ratio is $\frac{x \bmod m}{2R}$, where $x \in \mathbb{Z}/n\mathbb{Z}$ and R is the Montgomery constant. On the other hand, Walter and Thomson estimated that the ratio for a squaring is 0.33 and that for a multiplication is 0.25 if the modulus m is near to Montgomery constant R [14]. The attacker is able to distinguish a squaring and a multiplication by observing the final subtraction of Montgomery multiplication.

In this paper, we present some exact analysis on Montgomery multiplication under a reasonable assumption. Firstly we divide the interval between 0 and the underlying modulus into intervals with length R, then we investigate the distribution of outputs of multiplications in each interval. Integrating these results, we prove that the appearance ratios of the final subtraction in Montgomery multiplication and squaring are asymptotically $\frac{m}{4R}$ and $\frac{m}{3R}$, respectively. The assumption effects only the case that $m \approx R$ where m is the modulus and R is the Montgomery constant. Schindler's heuristic function $\frac{x \bmod m}{2R}$ is proved as well. This assumption describes clearly the behavior of the Montgomery multiplication's final subtractions.

We present a new invariant of the Montgomery multiplication as well. Namely we show that the multiplication $x \cdot y$ with $y = 3x \bmod m$ has a different subtraction ratio from both multiplication and squaring. This operation often appears in the addition formula of elliptic curve cryptosystems. We show a possible timing attack based on this invariant. Indeed, the randomization presented by Coron's 3rd [3] could be vulnerable to the attack. This is different to the attack of Goubin [5] because we have the opportunity of choosing in more points than a special one of the curve. Finally we show an experimental result on the appearance ratio discussed in this paper.

This paper is organized as follows: In Section 2 we shortly review the Montgomery multiplication and the timing attack using the final subtraction of the Montgomery multiplication. In Section 3 we present the proposed exact estimation about the appearance ratio of the final subtraction. In Section 4 we show a new timing attack and its analysis. In Section 5 we state the concluding remark.

2 Montgomery Multiplication and Timing Attack

In this section we shortly review the Montgomery multiplication and some timing attacks using the appearance probability of the last subtraction.

2.1 Montgomery Multiplication

The Montgomery Multiplication [9] is an efficient algorithm for computing modular multiplications without using relatively expensive divisions, and is widely utilized for public-key cryptosystems. Especially, it is suitable for the memory-constraint devices such as smart cards.

Note that the Montgomery multiplication has outputs slightly different from ordinary modular multiplications. In an exponentiation these can be corrected by three extra Montgomery multiplication. Because the Montgomery multiplication outputs results in the residue class without any divisions it is the fastest way to multiply. This is because if the radix b is chosen suitably the divisions are only shifts. Shifts are basic operations in hardware and are therefore fast.

The following algorithm is taken from [8]

ALGORITHM 1: MONTGOMERY MULTIPLICATION

Input: $m = (m_{n-1} \cdots m_0)_b,\ X = (x_{n-1} \cdots x_0)_b,\ Y = (y_{n-1} \cdot\ y_0)_b,\ b = 2^k,$
$R = b^n,\ \gcd(m, b) = 1,\ m' = -m^{-1} \bmod b.$
Output: $XYR^{-1} \bmod m$.
1. $A \leftarrow 0$ $(A = (a_n \cdots a_0)_b)$.
2. For i from 0 to $(n-1)$ do:
$temp \leftarrow 0$,
For j from 0 to $(n-1)$ do:
$\{temp, a_j\} \leftarrow x_j y_i + a_j + temp$,
$a_n \leftarrow temp,\ temp \leftarrow 0,\ u_i \leftarrow a_0 m' \bmod b$,
For j from 0 to n do:
$\{temp, a_j\} \leftarrow m_j u_i + a_j + temp$,
$A \leftarrow A/b$.
3. If $A \geq m$, $A \leftarrow A - m$. $\Leftarrow$ **Final Subtraction**
4. Return(A).

The running time of the steps can be analyzed as follows: The computations in step 2 are expected to take approximately constant time. This is because of the repetition in every multiplication and the constant n reputations of the for-loops. After step 2 the value of A varies between 0 and twice the modulus. A subtraction has to be done if A is larger than the modulus. This subtraction is called *final subtraction*.

2.2 Timing Attack and its Analysis

We shortly review the timing attack on RSA cryptosystem using the Montgomery multiplication.

Dhem et al. simulated a timing attack on the CASCADE smart card [4]. They focused that the probability of the final subtraction depends on the message and the secret bit. The attacker can guess the secret bit by observing the distribution of the final subtraction. The authors stated the final subtraction occurs in a multiplication of two random inputs in about 17% of the time. They expected a 512-bit RSA key to be cracked within a few minutes once 350 000 timing measurements are collected.

There are some theoretical estimations for the probability of the final subtraction. Walter and Thomson investigated the probability of the final subtraction appeared in Montgomery multiplication [14, 13, 11]. They showed the following estimations under several convenient conditions for simplicity.

$$P_{mul} = \frac{R}{4m}\left(1 - \left(1 - \frac{m}{R}\right)^2\right) - \left(1 - \frac{m}{R}\right) - \frac{R}{2m}\left(1 - \frac{m}{R}\right)^2 \log\left(1 - \frac{m}{R}\right), \quad (1)$$

$$P_{sqr} = 1 - \frac{2R}{3m}\left(1 - \left(1 - \frac{m}{R}\right)^{3/2}\right), \tag{2}$$

where P_{mul}, P_{sqr} are the probability of the final subtraction appeared in Montgomery multiplication for general multiplications and squarings respectively. Interestingly, the probability for squaring is $1/3$ and that for multiplication is $1/4$ for $m \approx R$. It is an open problem to show a general formula of the probability.

Schindler proposed another timing attack on RSA using the Chinese remainder theorem [10]. He estimated heuristically the probability of the final subtraction is

$$\frac{c \bmod m}{2R}, \tag{3}$$

where c is a ciphertext and m is the secret modulus. The secret modulus m can be calculated by the chosen ciphertext setting. As he stated in the paper, the precise proof for the formula is not given yet.

3 Exact Analysis of Montgomery Multiplication

In this section we analyze the distribution of the final subtraction in Montgomery multiplication. We will investigate the distribution for the general case and some special cases, and summarize these in section 3.5.

In case of $R = b$ ($n = 1$), Montgomery Multiplication is given by the following simple form.

ALGORITHM 2: MONTGOMERY MULTIPLICATION - SPECIAL CASE

Input: $m,\ X,\ Y,\ R$, $\gcd(m, R) = 1,\ m' = -m^{-1} \bmod R$
Output: $XYR^{-1} \bmod m$
S-1. $u \leftarrow xym' \bmod R$.
S-2. $A \leftarrow (xy + um)/R$.
S-3. If $A \geq m$, $A \leftarrow A - m$.
S-4. Return(A).

First of all, we will reduce the problem for ALGORITHM 1 to that for ALGORITHM 2 . Thus we will prove the following lemma.

Lemma 1. *For inputs of* ALGORITHM *1 and* ALGORITHM *2 , the final subtraction in step 3 of* ALGORITHM *1 is performed if and only if the final subtraction in step S-3 of* ALGORITHM *2 is performed.*

Proof. In step 2 of ALGORITHM 1 , in order to distinguish, let us denote A for each i by A_i. Then it can be easily seen that

$$A_{n-1} = \frac{xy + (\sum_{i=0}^{n-1} u_i)m}{b^n},$$

and we can see that the subtraction in step 3 is performed if and only if $A_{n-1} \geq mb^n = mR$. Let us set $S = \sum_{i=0}^{n-1} u_i$. Then by the validity of Montgomery

Multiplication, we have that A_{n-1} is an integer, namely, $xy + Sm \equiv 0 \bmod b^n$. Hence $S \equiv -xy/m \bmod R$. Moreover, as an integer, $S < R$, thus we have

$$S = (-xy/m \bmod R).$$

Note that the right hand side is an integer not less than 0 and less than R. Therefore, we have that the subtraction in step 3 is performed if and only if $xy + (-xy/m \bmod R)m \geq mR$, and this condition is nothing less than the equivalent condition for the final subtraction in step S-3 of ALGORITHM 2 . □

3.1 Preparation

In the following, we will consider the problem for ALGORITHM 2 . After step S-2 we obtain the following equation:

$$A = (xy + (xym' \bmod R)m)/R \tag{4}$$

Thus we can see that

$$A \geq m \;\Leftrightarrow\; xy + (xym' \bmod R)m \geq mR. \tag{5}$$

Here we set $w = xy$, and consider the approximation of the following equation

$$g(m, R) := \#\left\{w \in \mathbb{Z} \mid 0 \leq w \leq (m-1)^2,\ w + (wm' \bmod R)m \geq mR\right\}. \tag{6}$$

When we represent $w = \eta + \xi R$, $0 \leq \eta < R$, $0 \leq \xi \leq (m-1)^2/R$, then the equation in the left side of (6) becomes

$$\eta + \xi R + (\eta m' \bmod R)m. \tag{7}$$

This number should be divisible by R, so that we can represent $(\eta m' \bmod R)m = -\eta + \pi R$ for some integer $\pi = \pi(\eta)$ depending on η. Moreover, we know $\pi \leq (R-1)(m+1)/R$ due to $0 \leq (\eta m' \bmod R)m \leq (R-1)m$. Therefore, if $m < R-1$ holds, then we obtain $0 \leq \pi \leq m-1$. Next, we assume the following distribution.

Assumption *DIS.* $\alpha := \eta m' \bmod R$ distributes in interval $0 \leq \alpha < R$ uniformly and randomly for R-fold different η.

We know that this assumption is adequate experientially[†]. From this assumption, we can see that π distributes in interval $0 \leq \pi < m$ uniformly and randomly for R-fold different η. Indeed, for $0 < \eta, \eta' < m$, it is easy to see that $\pi(\eta) = \pi(\eta')$ if and only if $\eta = \eta'$. Moreover the random distribution of $\eta m' \bmod R$ induces the random distribution of π. Hence we can see that one π corresponds to $R/(m+1)$-fold η on average, namely there is an $(R/(m+1))$-to-1 map between π and η. On the other hand, Equation (7) can be represented as $R(\xi + \pi)$. If $\xi + \pi \geq m$, then $R(\xi + \pi)$ is greater than mR. For fixed ξ, the

[†] Since $m' \in (\mathbb{Z}/R\mathbb{Z})^\times$, the m'-multiplication map $\eta \mapsto \eta m' \bmod R$ is bijective. Thus $\eta m' \bmod R$ are uniformly distributed.

conditions in (6) is true with ξ-fold π that satisfies $m-\xi \le \pi \le m-1$. We know that ξ satisfies $0 \le \xi \le (m-1)^2/R$, and thus we have obtained

$$g(m,R) \;\approx\; \sum_{\xi=0}^{\frac{(m-1)^2}{R}} \frac{R}{m+1}\xi \;\approx\; \int_1^{\frac{(m-1)^2}{R}} \frac{Rx}{m+1}dx \;\approx\; \frac{m^3}{2R} - \frac{R}{2m}. \tag{8}$$

Note that we used $m \pm 1 \approx m$ for the final approximation.

3.2 Distribution of the Final Subtraction in The General Case

Next, we consider the distribution of $xyR^{-1} \bmod m$ with the final subtraction in the following. Previously we set $w = xy$, but xy is not uniformly distributed in interval $[0,(m-1)^2]$ for $0 \le x, y \le m-1$. We consider the divided interval $[0,(m-1)^2]$ with width R. In general, we set $G_N := \{0,1,2,\cdots,N-1\} \subset \mathbb{Z}$ for natural integer N and let $\phi = \phi_N$ be the multiplication map:

$$\phi : G_N \times G_N \to G_{(N-1)^2+1}, \quad (x,y) \mapsto xy.$$

For fixed ξ, the value $w = \eta + \xi R$ runs between ξR and $(\xi+1)R$. Denote by $F_\phi(\xi)$ the number of the images of $\phi_m : F_\phi(\xi) := \#\{\ \mathrm{Im}(\phi_m) \cap [\xi R, (\xi+1)R)\ \}$. Then, for fixed ξ the probability that the integers in $[\xi R, (\xi+1)R)$ are equal to the image of map ϕ_m is given by $F_\phi(\xi)/R$. In the words, the number of π that are contained in the image of ϕ_m is given by

$$\frac{F_\phi(\xi)}{R}\xi. \tag{9}$$

On the other hand, let $G_\phi(\xi)$ denote the number of integers $0 \le x, y \le m-1$ whose images by $\phi = \phi_m$ are in the interval $[\xi R, (\xi+1)R)$:

$$G_\phi(\xi) := \#\{(x,y) \in G_m \times G_m \mid \phi_m(x,y) \in [\xi R, (\xi+1)R)\}.$$

From $\xi R \le xy \le (\xi+1)R$ and the condition of x, y, we have

$$\frac{\xi R}{m} \le x < m, \tag{10}$$

and for a fixed x, the number of y that satisfies the conditions is exactly R/x (more precisely we should consider its floor value). Hence we have

$$G_\phi(\xi) \approx \sum_{\xi R/m \le x < m} \frac{R}{x} \approx R(2\log m - \log R - \log \xi).$$

Therefore, among the image of ϕ_m from the interval $[\xi R, (\xi+1)R)$, there are $G_\phi(\xi)/F_\phi(\xi) \approx R(2\log m - \log R - \log \xi)/F_\phi(\xi)$ elements mapped from (x,y) on average. Consequently, for fixed ξ, the number of images of the map ϕ_m is equal to $F_\phi(\xi)\xi/R$ among ξ-fold π. Each image has $R(2\log m - \log R - \log \xi)/F_\phi(\xi)$-fold

pre-images of (x, y) on average. Therefore, for w in $[\xi R, (\xi+1)R)$, the number that stratifies (4) with x, y is

$$\frac{R(2\log m - \log R - \log \xi)}{F_\phi(\xi)} \cdot \frac{F_\phi(\xi)}{R}\xi \cdot \frac{R}{m+1} = \frac{R(2\log m - \log R - \log \xi)\xi}{m+1}.$$

Let $t(m, R)$ denote the number of (x, y) that satisfies Equation (5):

$$s(m, R) := \#\left\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid 0 \le x, y \le m-1,\ xy + (xym' \bmod R)m \ge mR\right\}.$$

Then from the above argument, we have the following approximation formula.

$$\begin{aligned} s(m, R) &\approx \frac{R}{m+1} \sum_{\xi=1}^{(m-1)^2/R} \left(\log \frac{m^2}{R} - \log \xi\right) \xi \\ &\approx \frac{R}{m+1} \int_1^{(m-1)^2/R} \left(\log \frac{m^2}{R} - \log x\right) x dx \qquad (11) \\ &\approx \frac{R}{m+1} \left\{ \frac{1}{4}\left(\frac{(m-1)^2}{R}\right)^2 + \left(1 - \log \frac{m^2}{R}\right)\right\} \qquad (12) \\ &\approx \frac{m^3}{4R} + \frac{R}{m}\left(1 - \log \frac{m^2}{R}\right). \qquad (13) \end{aligned}$$

Here, the transformation from (11) to (13) is obtained by the partial derivation and $m \pm 1 \approx m$.

3.3 The Case of $x = y$

We consider the case of $x = y$, thus we will estimate the following.

$$t(m, R) := \#\left\{x \in \mathbb{Z} \mid 0 \le x \le m-1,\ x^2 + (x^2 m' \bmod R)m \ge mR\right\}.$$

We follow the estimation for the general case. Let $G_\psi(\xi)$ denote the number of integers $0 \le x \le m-1$ whose images by $\psi(x) = \psi_m(x) := x^2$ are in the interval $[\xi R, (\xi+1)R)$:

$$G_\psi(\xi) := \#\{x \in G_m \mid \psi_m(x) \in [\xi R, (\xi+1)R)\}.$$

Because of $\sqrt{\xi R} \le x < \sqrt{(\xi+1)R} < m$, we have

$$G_\psi(\xi) \ \approx \sum_{\sqrt{\xi R} \le x < \sqrt{(\xi+1)R}} 1 \ \approx\ \sqrt{(\xi+1)R} - \sqrt{\xi R}.$$

Hence, among the image of ψ_m in the interval $[\xi R, (\xi+1)R)$, there are $G_\psi(\xi)/F_\psi(\xi) \approx (\sqrt{(\xi+1)R} - \sqrt{\xi R})/F_\psi(\xi)$ elements mapped from x on average, where $F_\psi(\xi)$ denote the number of the images of $\psi_m : F_\psi(\xi) := \#\{\ \mathrm{Im}(\psi_m) \cap [\xi R, (\xi+1)R)\ \}$. Therefore, for w in $[\xi R, (\xi+1)R)$, the number that stratifies (5) with x is

$$\frac{\sqrt{(\xi+1)R} - \sqrt{\xi R}}{F_\psi(\xi)} \cdot \frac{F_\psi(\xi)}{R}\xi \cdot \frac{R}{m+1} = \frac{\sqrt{R}}{m+1}\left(\sqrt{(\xi+1)} - \sqrt{\xi}\right)\xi.$$

Thus we have following approximation.

$$\begin{aligned} t(m,R) &\approx \frac{\sqrt{R}}{m+1} \sum_{\xi=1}^{\frac{(m-1)^2}{R}} \left(\sqrt{\xi+1}-\sqrt{\xi}\right)\xi \\ &\approx \frac{\sqrt{R}}{m+1} \int_1^{\frac{(m-1)^2}{R}} \left(\sqrt{x+1}-\sqrt{x}\right) x dx \\ &\approx \frac{\sqrt{R}}{m+1} \left(\frac{1}{3}\left(\frac{(m-1)^2}{R}\right)^{3/2} + \frac{15}{8}\left(\frac{(m-1)^2}{R}\right)^{1/2} \right). \end{aligned}$$

As in the previous section, using $m \pm 1 \approx m$ and ignoring small constant, we have

$$t(m,R) \approx \frac{m^2}{3R}. \tag{14}$$

3.4 The Case of fixed x

We consider the case that x is fixed in the following. Let x be an integer such that $0 \leq x < m$, and fix. If the multiplication xy for $0 \leq y < m$ lies in the interval $[\xi R, (\xi+1)R)$, then from the equation (10), we have

$$\xi \leq \frac{mx}{R}. \tag{15}$$

In this case, for R/x-folds y, the image of ϕ_m is in $[\xi R, (\xi+1)R)$ (if $\xi > mx/R$, then no image for y is in this interval). On the other hand, we have to consider the distribution of $xym' \bmod R$ for m-fold y instead of that of $\eta m' \bmod R$ for R-fold η in Assumption *DIS*, and the former strongly depends on the fixed x. We will focus on the gcd of x and R in the following.

Lemma 2. *Let $x' = \gcd(x, R)$. Then for any $r(< R)$, there exists some $s = s(r) < R/x'$ such that $xr \bmod R = sx'$ ($< R$ as an integer).*

Proof. As an integer, let $xr = \alpha R + \beta$, $\beta \leq R-1$, then we have $\beta \equiv 0 \bmod x'$. Hence putting $\beta = sx'$ as an integer, we have $s \leq (R-1)/x'$ and $xr \bmod R = sx'$. □

Using this lemma, in the equation (4), there exists $s \leq (R/x') - 1$ such that $xym' \bmod R = x's$. Hence we have $xy + (xym' \bmod R)m = xy + x'sm \leq mR + xy - x'm$. Therefore, for y such that $y < x'm/x$, the subtraction is not performed. So from equation $\xi m/x < y$, for ξ satisfying

$$\xi < \frac{x'm}{R}, \tag{16}$$

the subtraction is not performed. Hence, similarly to the general case, an approximation of the number

$$u(x,m,R) := \#\{y \in \mathbb{Z} \mid 0 \leq y \leq m-1,\ xy + (xym' \bmod R)m \geq mR\}$$

is given by following (using $m \pm 1 \approx m$).

$$u(x,m,R) \;\approx\; \frac{R}{x(m+1)} \sum_{\xi=\frac{m\gcd(x,R)}{R}}^{\frac{mx}{R}} \xi \;\approx\; \frac{m}{2xR}(x^2 - \gcd(x,R)^2). \tag{17}$$

Remark 1. In case of $z := -x/m \bmod R$ is very small (e.g. $z = 1, 2, \ldots$) or very large(e.g. $z = R-1, R-2, \ldots$), we can see that there are some bias. In order to explain these bias, we need to consider u as a function of x, m, R and z.

3.5 Comparison of Probability

There are m^2 inputs for the general case and m inputs for the case of $x = y$, $y = ax \bmod m$ and fixed x. Therefore, from the previous sections, we have obtained the following probabilities.

$$\frac{g(m,R)}{m^2} \approx \frac{m}{2R}, \qquad \frac{s(m,R)}{m^2} \approx \frac{m}{4R}, \qquad \frac{t(m,R)}{m} \approx \frac{m}{3R},$$

$$\frac{u(x,m,R)}{m} \;\approx\; \frac{1}{2xR}(x^2 - \gcd(x,R)^2).$$

Consequently, we obtain the following theorem.

Theorem 1. *We assume that the assumption DIS is true. The final subtraction of Montgomery multiplication asymptotically appears with probability $\frac{m}{4R}$. If two inputs are equal (i.e. Montgomery squaring), then the probability becomes $\frac{m}{3R}$.*

If we choose $m \to R$, then these ratios for Montgomery multiplication and squaring converge $\frac{1}{4}$ and $\frac{1}{3}$, respectively. On the other hand, for $m \to R/2$, these ratios converge $\frac{1}{8}$ and $\frac{1}{6}$, respectively.

Corollary 1. *For randomly chosen m, the average ratio of the final subtraction in Montgomery multiplication (or Montgomery squaring) is asymptotically about 0.188 (or 0.250), respectively.*

Proof. From the assumption, m randomly distributes in interval $[\frac{R}{2}, R]$. Then the average ratio for Montgomery multiplication is $\frac{3}{16} = 0.1875$. Similarly, we can estimate $\frac{1}{4} = 0.25$ for Montgomery squaring. □

4 Application to Elliptic Curve Cryptosystems

In this section we shortly review elliptic curve cryptosystems, and side channel attack on them. Then we show a new invariant of a special Montgomery multiplication used for elliptic curve cryptosystems.

4.1 Elliptic Curve Cryptosystems

Elliptic curves over a finite prime field $K = GF(m)$ with $m > 3$ are defined by

$$E : \{(x, y) \in K^2 | y = x^3 + ax + b\} \cup \{\mathcal{O}\}, \tag{18}$$

where $a, b \in K,\ 4a^3 + 27b^2 \neq 0$, and $\mathcal{O}$ is a point at infinity. The Elliptic curve E has a group structure with neutral element $\mathcal{O}$. The group operation of the elliptic curve is as follows:

> Let E denote an elliptic curve and $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ denote points on the curve then $-P_1 = (x_1, -y_1)$. $P_3 = P_1 + P_2$ is calculated by
>
> $$x_3 = \begin{cases} \lambda_1^2 - x_1 - x_2 & : \quad P_1 \neq P_2 \\ \lambda_2^2 - 2x_1 & : \quad P_1 = P_2 \end{cases} \quad y_3 = \begin{cases} (x_1 - x_3)\lambda_1 - y_1 & : \quad P_1 \neq P_2 \\ (x_1 - x_3)\lambda_2 - y_1 & : \quad P_1 = P_2 \end{cases}$$
>
> where $\lambda_1 = \frac{y_1 - y_2}{x_1 - x_2}$ and $\lambda_2 = \frac{3x_1^2 + a}{2y_1}$.

We denote by ECADD by the first formula and ECDBL by the second, respectively. In order to avoid the expensive inversion operation in the affine coordinates, we usually use the *Jacobian* coordinates [2]. A point $P = (x, y)$ in the affine coordinates is represented by $P = (X, Y, Z)$ with $x = X/Z^2$ and $y = Y/Z^3$ in the Jacobian coordinates. The addition formula in the Jacobian coordinates is as follows:

ECDBL in Jacobian Coordinates (ECDBL$^{\mathcal{J}}$) :
$X_3 = T$, $Y_3 = -8{Y_1}^4 + M(S - T)$, $Z_3 = 2Y_1Z_1$,
$S = 4X_1{Y_1}^2$, $M = 3{X_1}^2 + a{Z_1}^4$, $T = -2S + M^2$.

ECADD in Jacobian Coordinates (ECADD$^{\mathcal{J}}$) :
$X_3 = -H^3 - 2U_1H^2 + R^2$, $Y_3 = -S_1H^3 + R(U_1H^2 - X_3)$, $Z_3 = Z_1Z_2H$,
$U_1 = X_1{Z_2}^2$, $U_2 = X_2{Z_1}^2$, $S_1 = Y_1{Z_2}^3$, $S_2 = Y_2{Z_1}^3$, $H = U_2 - U_1$, $R = S_2 - S_1$.

The group offers the scalar multiplication of $k \cdot P, k \leq ord(E)$ for a point P with order q on a curve E. A standard double-and-multiply algorithm can compute the scalar multiplication, but it is not secure against the timing attack. The double-and-multiply-always method can resist the timing attack [3].

Algorithm 3: Coron dummy

Input: $d = (d_{n-1} \cdots d_1 d_0)_2,\ P \in E(K)\ \ (d_{n-1} = 1)$
Output: dP.
1. $Q[0] \leftarrow P$
2. For $i = (n-2)$ down to 0 do:
   $Q[0] \leftarrow ECDBL(Q[0])$,
   $Q[0] \leftarrow ECADD(Q[0], P)$
   $Q[0] \leftarrow Q[d_i]$
3. Return($Q[0]$).

4.2 DPA and Countermeasure

The differential power analysis (DPA) observes many power consumptions and analyze these information together with statistic tools. Even if a method is secure against the timing attack, it might not be secure against the DPA. The DPA attacker tries to guess that the computation cP for an integer c is performed during the exponentiation. She gathers many power consumptions cP_i with $i \in 1, 2, 3, \ldots$, and detects the spike arisen from the correlation function based on the specific bit of cP_i. The DPA can break Algorithm 2, because the sequence of generated points is deterministic and the DPA can find correlations for a specific bit.

Coron pointed out that it is necessary to insert random numbers during the computation of dP to prevent DPA [3]. The randomization eliminates the correlation between the secret bit and the sequence of points. The main idea of these countermeasures is to randomize the base point before starting the scalar multiplication. If the base point is randomized, there is no correlation among the power consumptions of each scalar multiplication. The DPA cannot obtain the spike of the power consumption derived from the statistical tool. This countermeasure is based on randomization of Jacobian coordinates. To prevent DPA we transform $P = (x, y)$ in affine coordinate to $P = (r^2x, r^3y, r)$ in Jacobian coordinates for a random value $r \in K^*$. This randomization produces the randomization in each representation of the point and the randomization of power consumptions during scalar multiplication dP.

However, Goubin proposed a DPA on Coron's randomization [5]. He pointed out that the point $(0, y)$ can not be randomized by Coron's randomization. Akishita and Takagi extended his attack to the case of auxiliary registers, called zero-value point attack [1]. The attack adaptively chooses a base point P and observes side channel information of the scalar multiplication dP, where d is a secret scalar. The bits of the secret scalar can be recovered if the point $(0, y)$ or zero-valued register appears. For example, the second most bit d_{n-1} should be 1 in Algorithm 3 if and only if for the point $(0, y)$ appears during the scalar multiplication dP with base point $P = (6^{-1}\#E)(0, y)$.

4.3 Proposed Attack

We propose an new attack on Algorithm 3 using the Coron's 3rd randomization.

Recall that the recommended curve from SECG uses the curve coefficient $a = -3$ [12]. If a is chosen as $a = -3$, the auxiliary parameter $M = 3X^2 + aZ^4$ of ECDBL in the Jacobian coordinate is computed by $M = 3(X + Z^2)(X - Z^2)$, and the computation time of ECDBL is reduced from $10M$ to $8M$, where M is the cost of a multiplication in K.

Assume that the underlying curve has the point P whose x-coordinate is equal to 2 (i.e., $(2, y)$). This point is randomized by the Coron's 3rd method: $(2r^2, r^3y, r)$ with a random element $r \in K$. Then the auxiliary parameter M takes value $3(2r^2 + r^2)(2r^2 - r^2) = 3(3r^2)(r^2)$. This means that ECDBL with input $(2, y)$ is not totally randomized by the Coron's 3rd method — there is an invariant of multiplication with the form $(3r^2)(r^2)$ under the Coron's 3rd randomization.

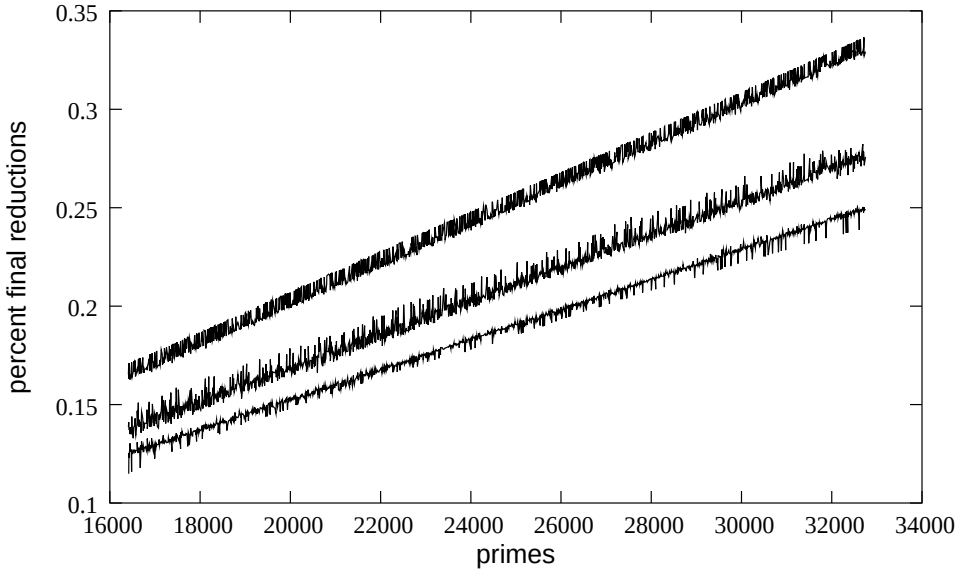

Fig. 1. The distribution of the final subtractions appeared in Montgomery multiplication for x^2 (upper), $x \cdot y$ with $y = 3 \cdot x \bmod m$ (middle), and $x \cdot y$ (lower)

Figure 1 shows that we can statistically distinguish the operation $xy \bmod m$ with $y = 3x \bmod m$ from other operations (e.g., multiplication or squaring). The lowest curve shows the percentage of final subtractions which take place in the computation of $x \cdot y$ with $0 \leq x, y < m$. The curve in the middle shows the results for $x \cdot y$ with $y = 3 \cdot x$ and the upper curve the results for x^2.

Because the distinction can be done a timing attack should be possible. The Coron's dummy method is vulnerable under the adaptive chosen ciphertext described in the previous section. We prove the distribution of the final subtraction appeared in $xy \bmod m$ with $y = 3x \bmod m$ in the following.

Theorem 2. *We assume that the assumption DIS is true. The final subtraction of Montgomery multiplication for xy with $y = 3x \bmod m$ asymptotically appears with probability $\frac{5m}{18R}$, where m is the underlying modulus.*

Proof. Let assume that $\gcd(3, m) = \gcd(3, R) = 1$ in this section. We consider the case of $y = 3x \bmod m$ in the following. Let $c(m, R)$ be the number of x that satisfies Equation (5):

$$c(m, R) := \#\left\{x \in \mathbb{Z} \mid 0 \leq x \leq m-1,\ y = 3x \bmod m,\ xy + (xym' \bmod R)m \geq mR\right\}.$$

We follow the estimation for the case of $x = y$. The number of integers $0 \leq x \leq m-1$ and whose images by $\phi_{m,3}(x) = x(3x \bmod m)$ are in the interval $[\xi R, (\xi+1)R)$ is

$$G_3(\xi) := \#\{x \in G_m \mid \phi_{m,3} \in [\xi R, (\xi+1)R)\}.$$

The function $\phi_{m,3}(x)$ is explicitly represented as follows:

$$\phi_{m,3}(x) = \begin{cases} 3x^2 & : \quad 0 \le x < \frac{m}{3} \\ x(3x-m) & : \quad \frac{m}{3} \le x < \frac{2m}{3} \\ x(3x-2m) & : \quad \frac{2m}{3} \le x < m. \end{cases}$$

Using the formula for solving quadratic equation, we can obtain the relationship:

$$G_3(\xi) \approx \begin{cases} \sqrt{\xi+1} - \sqrt{\xi} + \mu_1(\xi) - \mu_0(\xi) + \nu_1(\xi) - \nu_0(\xi) & : \quad 1 \le \xi < \frac{(m-1)^2}{3R} \\ \mu_1(\xi) - \mu_0(\xi) + \nu_1(\xi) - \nu_0(\xi) & : \quad \frac{(m-1)^2}{3R} \le \xi < \frac{2(m-1)^2}{3R} \\ \nu_1(\xi) - \nu_0(\xi) & : \quad \frac{2(m-1)^2}{3R} \le \xi < \frac{(m-1)^2}{3R}, \end{cases}$$

where $\mu_i(\xi) = \frac{\sqrt{m^2+12(\xi+i)R}}{6}$ and $\nu_i(\xi) = \frac{\sqrt{4m^2+12(\xi+i)R}}{6}$. From the same argument in the previous section, we are able to obtain the estimation about $c(m,R)$.

$$\begin{aligned} & c(m,R) \\ & \approx \frac{1}{m}\sqrt{\frac{R}{3}} \left(\sum_{\xi=1}^{\frac{(m-1)^2}{3R}} (\sqrt{x+1} - \sqrt{x}) + \sum_{\xi=1}^{\frac{2(m-1)^2}{3R}} (\mu_1(x) - \mu_0(x)) + \sum_{\xi=1}^{\frac{(m-1)^2}{R}} (\nu_1(x) - \nu_0(x)) \right) \\ & \approx \frac{1}{m}\sqrt{\frac{R}{3}} \left(\left(\frac{1}{3}\right)^{5/2} \left(\frac{m^2}{R}\right)^{3/2} + \frac{5\sqrt{3}}{54} \left(\frac{m^2}{R}\right)^{3/2} + \frac{4\sqrt{3}}{27} \left(\frac{m^2}{R}\right)^{3/2} \right) \\ & \approx \frac{5}{18}\frac{m^2}{R}. \end{aligned}$$

□

The average probability of occuring the final subtraction over randomly chosen K is $\frac{5}{24} = 0.208$, which is not equal to that of multiplication (0.188) or squaring (0.250). Similarly, we can prove that multiplication $x \cdot (ax)$ with small a has a different probability.

5 Conclusion

In this paper we presented some exact analysis related to the final subtraction of Montgomery multiplication. We investigated the distribution of outputs of multiplications in short intervals included between 0 and the underlying modulus. Integrating these results, we proved that the appearance ratios of the final subtraction during the Montgomery multiplication in the multiplication and squaring are asymptotically $\frac{m}{4R}$ and $\frac{m}{3R}$, respectively.

Based on the analysis we proposed a new invariant for the subtraction, namely multiplication $x \cdot (3x)$. We showed that this invariant appears at the randomization of parameter proposed by Coron, we could break it by DPA using the differences of the appearance ratios between general multiplications, squarings and the above case.

It is an interesting open problem to investigate further invariants of the Montgomery multiplication.

References

1. T. Akishita and T. Takagi, "Zero-Value Point Attacks on Elliptic Curve Cryptosystem", ISC 2003, LNCS 2851, pp.218-233, 2003.
2. H. Cohen, A. Miyaji, and T. Ono, "Efficient Elliptic Curve Exponentiation Using Mixed Coordinates", ASIACRYPT '98, LNCS 1514, pp. 51-65, 1998.
3. J.-S. Coron, "Resistance against Differential Power Analysis for Elliptic Curve Cryptosystems", CHES '99, LNCS 1717, pp. 292-302, 2002.
4. J.-F. Dhem, F. Koeune, P.-A. Leroux, P. Mestré, J.-J. Quisquater, and J.-L. Willems, "A Practical Implementation of the Timing Attack," CARDIS 1998, LNCS 1820, pp.167-182, 2002.
5. L. Goubin, "A Refined Power-Analysis Attack on Elliptic Curve Cryptosystems", PKC 2003, LNCS 2567, pp. 199-211, 2003.
6. C. Kocher, "Timing attacks on Implementations of Diffie-Hellman, RSA, DSS, and other Systems," CRYPTO '96, LNCS 1109, pp.104-113, 1996.
7. C. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," CRYPTO '99, LNCS 1666, pp.388-397, 1999.
8. A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1997.
9. P. L. Montgomery, "Speeding the Pollard and Elliptic Curve Methods of Factorization", Mathematics of Computation, vol. 48, pp. 243-264, 1987.
10. W. Schindler, "A Timing Attack against RSA with the Chinese Remainder Theorem," CHES 2000, LNCS 1965, pp.109-124, 2000.
11. W. Schindler and C. Walter, "More Detail for a Combined Timing and Power Attack against Implementations of RSA," IMA 2003, LNCS 2898, pp.245-263, 2003.
12. Standards for Efficient Cryptography Group (SECG), Specification of Standards for Efficient Cryptography. Available from `http://www.secg.org`
13. C. Walter, "Precise Bounds for Montgomery Modular Multiplication and Some Potentially Insecure RSA Moduli," CT-RSA 2002, LNCS 2271, pp.30-39, 2001.
14. C. Walter and S. Thompson, "Distinguishing Exponent Digits by Observing Modular Subtractions," CT-RSA 2001, LNCS 2020, pp.192-207, 2001.