ON THE UNIFORMITY OF DISTRIBUTION
OF THE RSA PAIRS

IGOR E. SHPARLINSKI

Abstract. Let $m = pl$ be a product of two distinct primes $p$ and $l$. We show that for almost all exponents $e$ with $\gcd(e, \varphi(m)) = 1$ the RSA pairs $(x, x^e)$ are uniformly distributed modulo $m$ when $x$ runs through

- the group of units $\mathbb{Z}_m^*$ modulo $m$ (that is, as in the classical RSA scheme);
- the set of $k$-products $x = a_{i_1} \cdots a_{i_k}$, $1 \leq i_1 < \cdots < i_k \leq n$, where $a_1, \ldots, a_n \in \mathbb{Z}_m^*$ are selected at random (that is, as in the recently introduced RSA scheme with precomputation).

These results are based on some new bounds of exponential sums.

1. Introduction

Let $m = pl$ be a product of two distinct primes $p$ and $l$, and let $\mathcal{E}_m$ be the set of integers $e$, $1 \leq e \leq \varphi(m)$, with $\gcd(e, \varphi(m)) = 1$, where $\varphi(N)$ is the Euler function. In this paper we consider the distribution modulo $m$ of the RSA pairs $(x, x^e)$. First of all we show that for almost all exponents $e \in \mathcal{E}_m$ this distribution is exponentially close to the uniform distribution, when $x \in U_m$ runs through the group of units $U_m = \mathbb{Z}_m^*$ modulo $m$. This result is an analogue of the results of [5, 6] about the uniformity of distribution of the Diffie–Hellman triples. Then we also consider the case when $x$ runs through all possible $k$-products of the form

$$x = \prod_{j=1}^{k} a_{i_j}, \quad 1 \leq i_1 < \cdots < i_k \leq n,$$

for some fixed $a_1, \ldots, a_n \in U_m$. We show that for almost all $n$-element sequences $a_1, \ldots, a_n \in U_m$, the corresponding pairs $(x, x^e)$ are uniformly distributed modulo $m$. Such pairs have been considered in [4] and provide a promising way to speed up the RSA encryption with precomputation.

Such uniformity of distribution results, although they do not have immediate security implications, still provide some useful information about pseudorandomness of the mapping $x \rightarrow x^e$, see [16]. In particular, it would be disastrous to discover that these pairs are not uniformly distributed; in this case one could guess their left-most bits with higher than average probability. Several other results about the uniformity of distribution and other properties of some pseudorandom generators of cryptographic interest are given in [9, 10, 11, 13, 24] for the power generator, which includes the RSA generator and the Blum–Blum–Shub generator.
(see [3, 7, 15, 17, 26]), and in [2, 12, 22, 23, 25] for the Naor–Reingold generator (see [18]).

As in [5, 6] our main tool is exponential sums. In fact our results directly depend on some estimates of these papers.

Throughout the paper all implicit constants in symbols “O” are absolute.

2. Notation and auxiliary results

Given a set $\mathcal{M}$ of $N$ points $(u_\nu, v_\nu) \in [0, 1]^2$, $\nu = 1, \ldots, N$, of the unit square, we define the discrepancy $D(\mathcal{M})$ of this set as

$$D(\mathcal{M}) = \sup_B \left| \frac{A_N(B)}{N} - \mu(B) \right|,$$

where the supremum is taken over all boxes $B = [\alpha, \beta] \times [\gamma, \delta] \in [0, 1]^2$, $\mu(B) = (\beta - \alpha)(\delta - \gamma)$ and $A_N(B)$ is the number of points of this set which hit $B$.

According to a standard principle, we can bound the discrepancy $D(\mathcal{M})$ by bounding the corresponding exponential sums. For arbitrary sets such a relation is given by the Erdős–Turán–Koksma inequality (see Theorem 1.21 of [8]) which we present in the following implicit form.

For an integer $a$ we define $\overline{a} = \max\{|a|, 1\}$.

**Lemma 1.** There exists an absolute constant $C > 0$ such that for any integer $L \geq 1$ the bound

$$D(\mathcal{M}) \leq C \left( \frac{1}{L} + \frac{1}{N} \sum_{0 < |r| + |s| < L} \frac{1}{r^2 s^2} \left| \sum_{\nu = 1}^{N} \exp(2\pi i (ru_\nu + sv_\nu)) \right| \right)$$

holds.

Let us define

$$e_d(z) = \exp(2\pi iz/d).$$

The following lemma shows how to reduce general exponential sums to exponential sums with prime power denominators (for example, see Problem 12.d to Chapter 3 of [27]).

**Lemma 2.** Let $m = m_1 m_2$, where $m_1 \geq m_2 \geq 2$ and $\gcd(m_1, m_2) = 1$, and let $k_1, k_2$ be such that

$$k_1 m_2 \equiv 1 \pmod{m_1} \quad \text{and} \quad k_2 m_1 \equiv 1 \pmod{m_2}.$$

Then for any polynomial $f(x)$ with integer coefficients

$$\sum_{x \in \mathcal{U}_m} e_m(f(x)) = \sum_{x_1 \in \mathcal{U}_{m_1}} e_{m_1}(k_1 f(x_1)) \sum_{x_2 \in \mathcal{U}_{m_2}} e_{m_2}(k_2 f(x_2)),$$

where $\mathcal{U}_m, \mathcal{U}_{m_1}$ and $\mathcal{U}_{m_2}$ are the groups of units modulo $m, m_1$ and $m_2$, respectively.

Indeed, this statement follows from Problem 12.d to Chapter 3 of [27] if one remarks that

$$k_1 m_2 + k_2 m_1 \equiv 1 \pmod{m}.$$

We also need an upper bound of certain double sums which is essentially the main result of [5].
Lemma 3. For any prime number $p$ the bound
\[ \max_{\gcd(r,s,p)=1} \left| \sum_{e=1}^{p-1} \sum_{x=1}^{p-1} e_p (rx + sx^e) \right|^4 = O(p^{14/3}) \]
holds.

Proof. Let $g$ be a primitive root modulo $p$. Then
\[ \sum_{y=1}^{p-1} \sum_{x=1}^{p-1} e_p (rx + sx^y)^4 = \sum_{y=1}^{p-1} \sum_{x=1}^{p-1} e_p (rg^x + sg^{xy})^4. \]
The last sum is estimated as $O(p^{14/3})$ (uniformly over all $r$ and $s$ with $\gcd(r, s, p) = 1$) in the proof of Theorem 8 of [5].

We define exponential sums
\[ W(r, s) = \sum_{e \in \mathcal{E}_m} \left| \sum_{x \in \mathcal{U}_m} e_m (rx + sx^e) \right|. \]

Lemma 4. Let $m = pl$, where $p$ and $l$ are two distinct primes. Then the bound
\[ \max_{\gcd(r,s,m)=1} W(r, s) = O \left( m^{23/12} \right) \]
holds.

Proof. Lemma 2 implies that there exist some integer numbers $k_p$ and $k_l$ with $\gcd(p, k_p) = \gcd(l, k_l) = 1$ and such that
\[ \sum_{x \in \mathcal{U}_m} e_m (rx + sx^e) = \sum_{x_1=1}^{p-1} e_p (k_p (rx_1 + sx_1^e)) \sum_{x_2=1}^{l-1} e_l (k_l (rx_2 + sx_2^e)). \]
From the previous equation and the Cauchy inequality we derive
\[
W(r, s) \leq \sum_{e \in \mathcal{E}_m} \sum_{x_1=1}^{p-1} e_p (k_p (rx_1 + sx_1^e)) \left| \sum_{x_2=1}^{l-1} e_l (k_l (rx_2 + sx_2^e)) \right| \\
\leq \varphi(m)^{1/2} \left( \sum_{e \in \mathcal{E}_m} \left| \sum_{x_1=1}^{p-1} e_p (k_p (rx_1 + sx_1^e)) \right|^4 \right)^{1/4} \\
\times \left( \sum_{e \in \mathcal{E}_m} \left| \sum_{x_2=1}^{l-1} e_l (k_l (rx_2 + sx_2^e)) \right|^4 \right)^{1/4} \\
\leq \varphi(m)^{1/2} \left( \frac{\varphi(m)}{p-1} \sum_{e=1}^{p-1} \left| \sum_{x_1=1}^{p-1} e_p (k_p (rx_1 + sx_1^e)) \right|^4 \right)^{1/4} \\
\times \left( \frac{\varphi(m)}{l-1} \sum_{e=1}^{l-1} \left| \sum_{x_2=1}^{l-1} e_l (k_l (rx_2 + sx_2^e)) \right|^4 \right)^{1/4}.
\]
Using the bound of Lemma 3, we obtain the desired result.
Therefore
\[ u \]

The contribution to this sum of each pair with
\[ u \]

Using the Cauchy inequality and changing the order of summation, we derive
holds.

Lemma 5. Let
\[ (2) \]

Let
\[ (1) \]

Thus
\[ (r,s,k) \]

For a given \( n \)-dimensional vector \( \mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{U}_m^n \) and a binary vector \( \mathbf{u} = (u_1, \ldots, u_n) \in \{0,1\}^n \) we put
\[ x_\mathbf{a}(\mathbf{u}) = \prod_{j=1}^n a_j^{u_j} \]

and define
\[ S_{k,n}(r,s) = \sum_{\mathbf{u} \in \mathcal{F}_{n,k}} \sum_{e \in \mathcal{E}_m} \left| \sum_{\mathbf{u} \in \mathcal{F}_{n,k}} e_m (r x_\mathbf{a}(\mathbf{u}) + s x_\mathbf{a}(\mathbf{u})) \right|. \]

Lemma 5. Let \( m = pl \), where \( p \) and \( l \) are two distinct primes. Then the bound
\[ \max_{\gcd(r,s,m) = 1} S_{k,n}(r,s) = O \left( m^{\frac{n}{2}} \left( |\mathcal{F}_{n,k}|^{1/2} + |\mathcal{F}_{n,k}|m^{-1/12} \right) \right) \]
holds.

Proof. Using the Cauchy inequality and changing the order of summation, we derive
\[ S_{k,n}(r,s)^2 \]
\[ \leq \left| \mathcal{U}_m^n \right| \left| \mathcal{E}_m \right| \sum_{\mathbf{u} \in \mathcal{U}_m^n} \sum_{e \in \mathcal{E}_m} \left| \sum_{\mathbf{u} \in \mathcal{F}_{n,k}} e_m (r x_\mathbf{a}(\mathbf{u}) + s x_\mathbf{a}(\mathbf{u})) \right|^2 \]
\[ = \left| \mathcal{U}_m^n \right| \left| \mathcal{E}_m \right| \sum_{\mathbf{u}, \mathbf{v} \in \mathcal{F}_{n,k}} \sum_{\mathbf{u} \in \mathcal{F}_{n,k}} \sum_{e \in \mathcal{E}_m} e_m (r x_\mathbf{a}(\mathbf{u}) + s x_\mathbf{a}(\mathbf{u}) - r x_\mathbf{a}(\mathbf{v}) - s x_\mathbf{a}(\mathbf{v})). \]

The contribution to this sum of each pair with \( \mathbf{u} = \mathbf{v} \) is \( \left| \mathcal{U}_m^n \right| \left| \mathcal{E}_m \right| \). For each pair \( \mathbf{u}, \mathbf{v} \in \mathcal{F}_{n,k} \) with \( \mathbf{u} \neq \mathbf{v} \) we can find \( i \) and \( j \), \( 1 \leq i < j \leq n \), with \( u_i = v_j = 1 \) and \( u_j = v_i = 0 \). Without loss of generality we may assume that \( i = 1, j = 2 \). In this case \( x_\mathbf{a}(\mathbf{u}) = Aa_1 \) and \( x_\mathbf{a}(\mathbf{v}) = Ba_2 \), where \( A \) and \( B \) do not depend on \( a_1 \) and \( a_2 \). Therefore
\[ \sum_{\mathbf{u} \in \mathcal{U}_m^n - 2} \sum_{e \in \mathcal{E}_m} e_m (r x_\mathbf{a}(\mathbf{u}) + s x_\mathbf{a}(\mathbf{u}) - r x_\mathbf{a}(\mathbf{v}) - s x_\mathbf{a}(\mathbf{v})) \]
\[ = \sum_{a_1, \ldots, a_n \in \mathcal{U}_m} \sum_{e \in \mathcal{E}_m} e_m (r A a_1^e + s A^e a_1) \sum_{a_2 \in \mathcal{U}_m} e_m (-r B a_2 - s B^e a_2), \]
where \( A \) and \( B \) depend only on \( u, v \) and \( a_3, \ldots, a_n \). Furthermore, by the Cauchy inequality we obtain

\[
\left| \sum_{e \in E, a_1 \in U_m} e_m (rAa_1^e + sA^e_1) \sum_{a_2 \in U_m} e_m (-rBa_2 - sB^e_2) \right|^2 \\
\leq \sum_{e \in E} \left| \sum_{a_1 \in U_m} e_m (rAa_1^e + sA^e_1) \right|^2 \times \left| \sum_{a_2 \in U_m} e_m (rBa_2 + sB^e_2) \right|^2.
\]

Taking into account that \( A, B \in U_m \), as in the proof of Lemma 4 we obtain that each factor in the above expression is \( O \left( m^{17/6} \right) \). We have

\[
|U_m| = \varphi(m) = (p-1)(l-1) \geq 0.25m.
\]

Therefore \( m^{17/6} = O(|U_m|^2m^{5/6}) \) and the desired result follows.

As after Lemma 4 we also remark that

\[
\max_{\gcd(r,s,m) = p} S_{k,n}(r, s) = O \left( m|U_m|^n \left( |F_{n,k}|^{1/2} + |F_{n,k}|^{l^{-1/12}} \right) \right)
\]

and

\[
\max_{\gcd(r,s,m) = l} S_{k,n}(r, s) = O \left( m|U_m|^n \left( |F_{n,k}|^{1/2} + |F_{n,k}|^{p^{-1/12}} \right) \right).
\]

Finally we recall that there exists an absolute constant \( c > 0 \) such that the Euler function \( \varphi(N) \) satisfies the inequality

\[
\varphi(N) \geq c \frac{N}{\log \log N}
\]

for any integer \( N \geq 2 \), (for example, see Problem 9.g to Chapter 2 of [27]).

### 3. Distribution of the RSA pairs

Now we are prepared to formulate our main results.

Denote by \( D_e \) the discrepancy of the pairs of fractional parts

\[
\left( \left\{ \frac{x}{m} \right\}, \left\{ \frac{x^e}{m} \right\} \right), \quad x \in U_m.
\]

**Theorem 6.** Let \( m = pl \), where \( p \) and \( l \) are two distinct primes. Then the bound

\[
\frac{1}{|E_m|} \sum_{e \in E_m} D_e = O(m^{-1/12} \log^2 m \log \log m)
\]

holds.
Proof. Select \( L = m \). Combining Lemma 1 with Lemma 4 and the bounds (1) and (2), we derive

\[
\sum_{e \in \mathcal{E}_m} D_e = O \left( 1 + \sum_{\substack{0 < |r| + |s| < m \\ \gcd(r, s, m) = 1}} \frac{m^{11/12}}{\sqrt{s}} + \sum_{\substack{0 < |r| + |s| < m \\ \gcd(r, s, m) = p}} \frac{ml^{-1/12}}{\sqrt{s}} + \sum_{\substack{0 < |r| + |s| < m \\ \gcd(r, s, m) = l}} \frac{mp^{-1/12}}{\sqrt{s}} \right)
\]

\[
= O \left( 1 + m^{11/12} \log^2 m + mp^{-1/12} l^{-1/12} \log^2 l + ml^{-1/12} p^{-1/12} \log^2 p \right)
\]

\[
= O \left( m^{11/12} \log^2 m \right).
\]

Recalling that \(|\mathcal{E}_m| = \varphi(\varphi(m))\) and taking into account the bound (5) and the inequality \( \varphi(m) \geq 0.25m \), we obtain the desired result.

In particular, we see that for any \( \delta > 0 \) for a random exponent \( e \) chosen uniformly from \( \mathcal{E}_m \) with probability at least \( 1 - \delta \), the bound

\[
D_e = O \left( \delta^{-1} m^{-1/12} \log^2 m \log \log m \right)
\]

holds.

Given integers \( 1 \leq k \leq n \) and an \( n \)-dimensional vector \( \mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{U}_m^n \), denote by \( D_{\mathbf{a},k,e} \) the discrepancy of the pairs of fractional parts

\[
\left( \left\{ \frac{x_{\mathbf{a}}(\mathbf{u})}{m} \right\} \right), \left( \left\{ \frac{x_{\mathbf{a}}(\mathbf{u})}{m} \right\} \right), \quad \mathbf{u} = (u_1, \ldots, u_n) \in \mathcal{F}_{n,k},
\]

where

\[
x_{\mathbf{a}}(\mathbf{u}) = \prod_{j=1}^{n} a_j^{u_j}.
\]

Using Lemma 6 and the bounds (3) and (4), in the same way as we have used Lemma 3 and the bounds (1) and (2) in the proof of Theorem 6, we obtain the following statement.

**Theorem 7.** Let \( m = pl \), where \( p \) and \( l \) are two distinct primes. Then the bound

\[
\frac{1}{|\mathcal{U}_m|^n|\mathcal{E}_m|} \sum_{\mathbf{a} \in \mathcal{U}_m^n} \sum_{e \in \mathcal{E}_m} D_{\mathbf{a},k,e} = O \left( \left( |\mathcal{F}_{n,k}|^{-1/2} + m^{-1/12} \right) \log^2 m \log \log m \right)
\]

holds.

In particular, we see that for any \( \delta > 0 \) for a random vector \( \mathbf{a} \) and a random exponent \( e \) chosen uniformly and independently from \( \mathcal{U}_m^n \) and \( \mathcal{E}_m \) with probability at least \( 1 - \delta \) the bound

\[
D_{\mathbf{a},k,e} = O \left( \delta^{-1} \left( |\mathcal{F}_{n,k}|^{-1/2} + m^{-1/12} \right) \log^2 m \log \log m \right)
\]

holds.
4. Remarks

Let $p$ be a prime and let $g$ be an element of a finite field $\mathbb{F}_p$ of $p$ elements of multiplicative order $t$.

As we have mentioned, an analogue of Theorem 6 for Diffie–Hellman triples $(g^x, g^y, g^{xy})$ has been obtained in [5, 6] (provided that $t$ is large enough). On the other hand, obtaining an analogue of Theorem 7 is an interesting open problem which is related to the Diffie–Hellman scheme with precomputation. In particular, similar questions have been briefly addressed in [20, 21]. More specifically, we are interested in establishing the uniformity of distribution of the following pairs of fractional parts

$$\left( \left\{ \frac{z_b(u)}{t} \right\}, \left\{ \frac{g^{z_b(u)}}{p} \right\} \right), \quad u = (u_1, \ldots, u_n) \in \mathcal{F}_{n,k},$$

where

$$z_b(u) = \sum_{j=1}^{n} b_j u_j,$$

for a random $n$-dimensional vector $b = (b_1, \ldots, b_n) \in \mathbb{Z}_t^n$ over the residue ring modulo $t$.

Even studying the distribution of only the first component, that is, just vectors $z_b(u), u \in \mathcal{F}_{n,k}$, would be of interest, see [20, 21]. We remark that several uniformity of distribution results about the vectors $z_b(u)$, when $u$ runs through all $n$-dimensional binary vectors, are known [1, 2, 12, 14, 19, 21, 25] and have some cryptographic applications.

Acknowledgment

The author would like to thank Phong Nguyen for a number of fruitful discussions.

References


DEPARTMENT OF COMPUTING, MACQUARIE UNIVERSITY, SYDNEY, NEW SOUTH WALES 2109, AUSTRALIA

E-mail address: igor@ics.mq.edu.au