HAL
open science

Statistical Properties of Short RSA Distribution and Their Cryptographic Applications

Pierre-Alain Fouque, Jean-Christophe Zapalowicz

To cite this version:

Pierre-Alain Fouque, Jean-Christophe Zapalowicz. Statistical Properties of Short RSA Distribution and Their Cryptographic Applications. Computing and Combinatorics, Aug 2014, Atlanta, United States. pp.525 - 536, 10.1007/978-3-319-08783-2_45 . hal-01094059

HAL Id: hal-01094059

https://inria.hal.science/hal-01094059v1

Submitted on 11 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical properties of short RSA distribution and their cryptographic applications

Pierre-Alain Fouque[1] and Jean-Christophe Zapalowicz[2]

[1] Université de Rennes 1 and Institut Universitaire de France
Pierre-Alain.Fouque@univ-rennes1.fr
[2] Inria, jean-christophe.zapalowicz@inria.fr

Abstract. In this paper, we study some computational security assumptions involve in two cryptographic applications related to the RSA cryptosystem. To this end, we use exponential sums to bound the statistical distances between these distributions and the uniform distribution. We are interesting studying the k least (or most) significant bits of $x^e \bmod N$, where N is a RSA modulus when x is restricted to a small part of $[0, N)$. First of all, we provide the first rigorous evidence that the cryptographic pseudo-random generator proposed by Micali and Schnorr is based on firm foundations. This proof is missing in the original paper and do not cover the parameters chosen by the authors. Consequently, we extend the proof to get a new result closer to the parameters using a recent work of Wooley on exponential sums and we show some limitations of our technique. Finally, we look at the semantic security of the RSA padding scheme called PKCS#1 v1.5 which is still used a lot in practice. We show that parts of the ciphertexts are indistinguisable from uniform bitstrings.

Keywords: Exponential Sums, Security Proof for Micali-Schnorr pseudo-random generator, semantic security of RSA padding scheme

1 Introduction

The RSA assumption states that given a random value y in $\mathbb{Z}_N$, where N is the product of two large primes, it is difficult to compute a e-th root of y, *i.e.* find x such that $y = x^e \bmod N$. The RSA problem has been studied by mathematicians and no attack more efficient than factoring the RSA modulus has been found since its discovery. Usually, it is very difficult to prove a computational assumptions such as RSA and cryptographers try to prove that this assumption is at least as difficult than another one for instance factoring. However, some evidence for the non-equivalence of these two hard problems has been provided by Boneh and Venkatesan in [8] while the RSA assumption seems to hold.

RSA is a valid cryptographic assumption since on average its difficulty seems to be established thanks to its self-reducibility property. Indeed, it is well-known than if we are able to invert RSA on a non-negligible subset of $\mathbb{Z}_N$, then we can invert nearly all values in $\mathbb{Z}_N$ with high probability. Based on this assumption, cryptographers have proposed and proved that RSA signature and encryption schemes using specific padding function [5,4] are secure in the random oracle model [3]. The security proof of RSA-OAEP for encryption has been showed to be flawed since, and many reparations have been proposed in [19,13] and checked with computational proof assistant [1].

Another direction to assess the security of a computational assumption consists in showing that the values we are looking for are computationally or statistically indistinguishable from the uniform distribution on bitstrings of the same size. Consequently, the best the adversary can do is to guess this value until he finds it. For RSA, it is easy to see that the value x is uniformly distributed if y is. In this paper, we will be interested by the short RSA problem: given y find x such that $x^e \bmod N$ given the promise that $x < M \ll N$. Coppersmith results [9] show that if $M = N^{1/e}$, then it is possible in polynomial-time to recover all short values for x using lattice reduction technique. However, we can wonder what is the security of this new assumption when $M \gg N^{1/e}$. It is trivial to see that if $M = N^{\frac{1}{e}+\epsilon}$ then, by guessing the high order bits of x, in time N^{ϵ} times a polynomial in $\log N$, we can invert x. However for larger values of M, the problem seems to be hard. This assumption is made in some standards, such as the standard PKCS#1 v1.5 that is used to protect RSA encryption and we will study it here in a special case. To assess the security of this short RSA assumption, the classical technique consists in studying the distribution of values $x^e \bmod N$ when $x < M$. This distribution cannot be uniform in $\mathbb{Z}_N$ since the output is larger than M, but it can be computationally difficult to distinguish it from the uniform distribution in $\mathbb{Z}_N$. We will study the short RSA distribution when we consider only some part of the output bits. In this case, it is possible to prove some mathematical statement on the statistical distance between this distribution and the uniform distribution. We will also study the security of the Micali-Schnorr pseudorandom generator. At COCOON 2013, we show some attacks that explain the choice of the parameters proposed by Micali and Schnorr [11]. This generator assumes that the distribution of the least significant bits of $x^e \bmod N$

is indistinguishable from the uniform distribution in $\{0,1\}^k$ if $x < M$ with $M = N^{2/e}$.

Our contributions. In the first part of this paper, we will prove the following informal theorem for different values of M.

Theorem 1. *Let $N = pq$ the product of two large primes and $M < N$. Let the function $f : \mathbb{Z}_M \longrightarrow \mathbb{Z}_N$ defined as $f(x) = x^e \mod N$. The k least significant bits of $f(x)$ for $k < \log N$ are statistically indistinguishable from the uniform distribution on $\{0,1\}^k$.*

For $M \gg \sqrt{N}$, we will show it using classical bounds based on Polya-Vinogradov bounds, and for $N^{1/e} \ll M \ll \sqrt{N}$, we will use more recent results proved by Wooley [21]. This last bound is very close to be optimal since for $M \leq N^{1/e}$, it is possible in polynomial-time to recover $x \leq M$ given the $k \geq M$ least significant bits of $f(x)$ as we explain in [11]. Indeed, in this case the function f becomes non modular and the problem of retrieving x is quite easy by using Hensel's lifting.

In the second part of this paper, we will show two applications of this theorem. Micali and Schnorr original proof just refers to the first bound $M \gg \sqrt{N}$ but the proof is missing and they do not give any hint to explain their more aggressive choice of parameters. Indeed, it would be possible to output less bits at each iteration of the generator, but the efficiency of this generator would be less efficient than Blum-Blum-Shub generator [7] for instance. Micali and Schnorr prefer to output more bit and avoid the previous attack when $M = N^{1/e}$. Last year, we develop some attacks to go beyond this bound using some time/memory tradeoff techniques which require exponential time complexity. Here, we show that these attacks are also near to be optimal since we can prove the indistinguishability of the random strings. We also explain that the parameters proposed by Micali and Schnorr are closed to be optimal in the special case of $e = 3$.

Finally, we propose to study the semantic security of the RSA padding called PKCS #1 v1.5 proposed by the RSA Labs. This padding is used in practice in many IETF standards and its security has been studied in [2] under some security notions. Here, we study the semantic security, *i.e.* the most interesting security notion, to assess this assumption as much as we can using mathematical and rigorous statements. This notion means that no bit of the plaintext leaks when we see the ciphertext. In this paper, we will show that no bit of the plaintext leaks when we see some part of the bits of the ciphertext.

2 Some Mathematical Backgrounds

Statistical Distance. We first define as in [14] the statistical distance which allows to measure the distance between two probability distributions, and the notion of statistically indistinguishability. In a security point of view, if the statistical distance between two distributions is negligible, this means that even a powerful and unbounded adversary will not be able to distinguish them. We will denote by U_k the uniform distribution on k-bit strings.

Definition 1. *Let X_n and Y_n be two arbitrary probability ensembles over a finite set $\mathcal{X}$. We say that X_n and Y_n have statistical distance δ bounded by $\Delta(n)$ if the following holds:*

$$\delta = \sum_{x \in \mathcal{X}} \left| Prob_{X_n}[x] - Prob_{Y_n}[x] \right| \leq \Delta(n).$$

We say that X_n and Y_n are statistically indistinguishable if for every polynomial $P(\cdot)$ and for sufficiently large n, $\Delta(n) \leq 1/P(n)$.

In order to bound our desired statistical distance, we will consider the notion of collision probability defined as follows:

Definition 2. *Let X and X' be two independent and identically distributed random variables with values in a finite set $\mathcal{X}$. The collision probability of X, denoted by $\mathrm{Col}(X)$ is the probability $\Pr[X = X'] = \sum_{x \in \mathcal{X}} \Pr[X = x]^2$.*

Finally, one needs a link between the statistical distance and the collision probability. That is the lemma presented in [20]:

Lemma 1. *Let X be a random variable with values in a set $\mathcal{X}$ of size $|\mathcal{X}|$ and $\mathrm{SD}(X, U_{\mathcal{X}}) = \delta$ the statistical distance between X and $U_{\mathcal{X}}$, the uniformly distributed variable over $\mathcal{X}$. We have:*

$$2\delta \leq \sqrt{|\mathcal{X}| \, \mathrm{Col}(X) - 1}.$$

Exponential Sums. Our proofs will bring into play a well-known tool in analytical number theory, that is exponentials sums. Thus it is convenient to use the notation for the following character of $\mathbb{Z}_m$:

$$\forall x \in \mathbb{Z}_m, e_m(x) = e^{\frac{2i\pi x}{m}} \in \mathbb{C}^*.$$

and we recall a notable property about sums of characters:

Theorem 2.

$$\sum_{x=0}^{m-1} e_m(ax) = \begin{cases} 0, & \text{if } a \not\equiv 0 \bmod m, \\ m, & \text{if } a \equiv 0 \bmod m. \end{cases}$$

Weil bound [17] allows to bound the exponential sums of the evaluations of a polynomial in term of the degree and the square root of the field:

Lemma 2. *For any prime p and any polynomial $f(X) \in \mathbb{F}_p[X]$ of degree $d \geq 1$ which is not identical to a constant modulo p, the following bound holds:*

$$\left| \sum_{x=1}^{p} e_p(f(x)) \right| < dp^{1/2}.$$

The next lemma is useful to reduce exponentials sums modulo a composite integer N to exponentials sums modulo prime numbers p and q that are involved in the prime decomposition of $N = pq$:

Lemma 3. *Let $N = pq$ a RSA modulus with p, q two large prime numbers and let u, v be such that $uq = 1 \bmod p$ and $vp = 1 \bmod q$. Then, for any polynomial $f \in \mathbb{Z}[X]$ with integer coefficients:*

$$\sum_{j=0}^{N-1} e_N(f(j)) = \sum_{j_1=0}^{p-1} e_p(uf(j_1)) \sum_{j_2=0}^{q-1} e_q(vf(j_2)).$$

Proof in appendix A.1.

Finally let us denote the following exponential sum for any integer a such as $\gcd(a, N) = 1$ and $M < N$:

$$S(a, M) = \sum_{0 \le x < M} e_N(ax^e).$$

Bounding this sum for different values of M will be at the heart of our proof.

3 Main Results

In order to prove Theorem 1, we have to estimate the statistical distance between the function $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and the uniform distribution on $\{0,1\}^k$. Then, in function of the values of N, e, k and M, we will be able to show (or not) the statistically indistinguishability of these two probability ensembles. A preliminary and general lemma is considered:

Lemma 4. *Let $N = pq$ the product of two large primes and M an integer less than N. Let B a bound on $S(a, M)$ for any integer a such as $\gcd(a, N) = 1$. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and U_k is bounded by:*

$$\delta \le \sqrt{\frac{2^k}{N}} + \frac{2^{k/2} B \log^{1/2} N}{M}.$$

Proof. Let us denote by $X_{M,e,N,k}$ the random variable whose values are taken in $[0, 2^k)$ with the following distribution: x is chosen uniformly at random in $\mathbb{Z}_M$ and we output $f(x) = \mathsf{lsb}_k(x^e \bmod N)$. We are interested in bounding the collision probability of this random variable. To count the number of values x and y such that $x^e \bmod N$ and $y^e \bmod N$ share the same k least significant bits, we introduce the following characteristic function:

$$\mathbb{1}(x, y, u) = 1/N \times \sum_{a=0}^{N-1} e_N(a(x^e - y^e - 2^k \cdot u)),$$

where $\mathbb{1}(x, y, u)$ is equal to 1 if $x^e - y^e = 2^k u \bmod N$ and 0 otherwise. By denoting $K = \lfloor \frac{N-1}{2^k} \rfloor$, we can evaluate $\mathrm{Col}(X_{M,e,N,k})$:

$$\begin{aligned}\mathrm{Col}(X_{M,e,N,k}) &= \frac{1}{M^2} \times \left|\{(x,y) \in [0, M-1]^2 \mid \exists u \leq K, x^e - y^e = 2^k \cdot u \bmod N\}\right|, \\ &= \frac{1}{M^2 N} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} \sum_{u=0}^{K} \sum_{a=0}^{N-1} e_N(a(x^e - y^e - 2^k \cdot u)).\end{aligned}$$

We need to do some modifications to evaluate this collision probability. First note that

$$\sum_{x=0}^{M-1} \sum_{y=0}^{M-1} e_N(ax^e - ay^e) = S(a,M)\overline{S(a,M)} = |S(a,M)|^2.$$

We thus obtain by considering the case $a = 0$ and $a \neq 0$:

$$\mathrm{Col}(X_{M,e,N,k}) = \frac{K+1}{N} + \frac{1}{M^2 N} \sum_{a=1}^{N-1} |S(a,M)|^2 \left(\sum_{u=0}^{K} e_N(-a2^k \cdot u) \right).$$

We now focus on the term $\sum_{a=1}^{N-1}(\sum_{u=0}^{K} e_N(-a2^k \cdot u))$. We first make a change of variables ($a \in \mathbb{Z}_N \setminus \{0\} \to 2^k a \in \mathbb{Z}_N \setminus \{0\}$ being a bijection), then use the fact that the sum is a geometric sequence, and finally the inequality $sin(y) \geq 2y/\pi$ if $0 \leq y \leq \pi/2$. In others words:

$$\begin{aligned}\sum_{a=1}^{N-1} \left| \sum_{u=0}^{K} e_N(-a2^k \cdot u) \right| &= \sum_{a=1}^{N-1} \left| \sum_{u=0}^{K} e_N(-a \cdot u) \right| = \sum_{a=1}^{N-1} \left| \frac{1 - e_N(-a(K+1))}{1 - e_N(-a)} \right|, \\ &= \sum_{a=1}^{N-1} \left| \frac{\sin\left(\frac{\pi a(K+1)}{N}\right)}{\sin(\frac{\pi a}{N})} \right| = 2 \sum_{a=1}^{\frac{N-1}{2}} \left| \frac{\sin\left(\frac{\pi a(K+1)}{N}\right)}{\sin(\frac{\pi a}{N})} \right|, \\ &\leq 2 \sum_{a=1}^{\frac{N-1}{2}} \left| \frac{1}{\sin(\frac{\pi a}{N})} \right| \leq 2 \sum_{a=1}^{\frac{N-1}{2}} \left| \frac{N}{a} \right| \leq N \log N.\end{aligned}$$

This is a well-known result first proved by Polya and Vinogradov. If we bound the last term uniformly in a $\left|S(a,M)\right|^2$ by B^2, we can finally bound $\mathrm{Col}(X_{M,e,N,k})$ using the triangular inequality:

$$\begin{aligned}\mathrm{Col}(X_{M,e,N,k}) &\leq \frac{K+1}{N} + \frac{1}{M^2 N} \cdot N \log N \cdot B^2, \\ &\leq \frac{K+1}{N} + \frac{B^2 \log N}{M^2}.\end{aligned}$$

We now use Lemma 1 to bound the statistical distance δ with the collision probability:

$$2\delta \leq \sqrt{\frac{2^k(K+1)}{N} - 1} + \frac{2^{k/2} B \log^{1/2} N}{M}.$$

Note to conclude that $\left|\frac{K+1}{N} - 1/2^k\right| \leq 1/N$. Indeed:

$$-\frac{1}{N} \leq \frac{N-1}{2^k} \cdot \frac{1}{N} - \frac{1}{2^k} \leq \left(1 + \left\lfloor \frac{N-1}{2^k} \right\rfloor\right) \cdot \frac{1}{N} - \frac{1}{2^k} \leq \frac{1}{N} + \frac{N-1}{2^k} \cdot \frac{1}{N} - \frac{1}{2^k} \leq \frac{1}{N}.$$

and consequently (with an omission of a factor 2 for sake of simplicity),

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2} B \log^{1/2} N}{M}.$$

□

Now that we have this technical lemma, it remains to evaluate the bound B of $S(a, M)$. Depending on the value of M, this bound will be different. More precisely, we propose a first evaluation where the statistical distance is negligible when M is greater than $\sqrt{N}$ (see Theorem 3) and a second one with any requirement on M (see Theorem 4). However this second evaluation is clearly bad compared to the first one in the case where M is greater than $\sqrt{N}$, and thus is interesting only if M is relatively small.

3.1 First Bound when $M \gg \sqrt{N}$

The first method to evaluate B proposes to bound an incomplete exponential sum given the bound of the complete one using another result of Polya and Vinogradov. We obtain a bound on δ which will be small only if the parameter M is sufficiently larger than $\sqrt{N}$:

Theorem 3. *Let $N = pq$ the product of two large primes and M an integer such as $M \gg \sqrt{N}$. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and U_k is negligible:*

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2} e^2 \sqrt{N} \log^{3/2} N}{M}.$$

Proof. In order to prove this theorem, we need the following lemma that we prove in Appendix A.2:

Lemma 5. *If for all $a \in \mathbb{Z}^*, b \in \mathbb{Z}$ and for any polynomial $P \in \mathbb{Z}[X]$,*

$$\left| \sum_{x \in \mathbb{Z}_N} e_N(aP(x)) e_N(bx) \right| \leq C,$$

then, for $0 \leq M \leq N$,

$$\left| \sum_{1 \leq x \leq M} e_N(aP(x)) \right| \leq C \log N.$$

From the notations of Lemma 5, we put $P(x) = x^e$ and $b = 0$ and we evaluate $\left|\sum_{x\in\mathbb{Z}_N} e_N(ax^e)\right|$ using Lemma 3 by denoting $u, v \in \mathbb{Z}$ such as $up + vq = 1$:

$$\left|\sum_{x\in\mathbb{Z}_N} e_N(ax^e)\right| = \left|\sum_{x\in\mathbb{F}_p} e_p(v \times ax^e) \sum_{x\in\mathbb{F}_q} e_q(u \times ax^e)\right|.$$

Then we consider the Weil bound (see Lemma 2):

$$\left|\sum_{x\in\mathbb{Z}_N} e_N(ax^e)\right| \leq ep^{1/2} \times eq^{1/2} \leq e^2 N^{1/2},$$

and finally Lemma 5 gives us the bound on $|S(a, M)|$, i.e:

$$|S(a, M)| \leq e^2 N^{1/2} \log N.$$

The replacement of B in Lemma 4 with this evaluation concludes the proof.

□

3.2 Second Bound when $M \ll \sqrt{N}$

It remains to treat the case where M is smaller than $\sqrt{N}$, specially if one wants to approach the optimal bound, i.e. $N^{1/e}$. Even if the following lemma and corollaries do not require anything on the size of M (except that it is less than N obviously), the bounds we find for δ are only interesting for small values of M, meaning $M \ll \sqrt{N}$.

Theorem 4. *Let $N = pq$ the product of two large primes and M an integer. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and U_k is:*

$$\delta \leq \sqrt{\frac{2^k}{N}} + 2^{k/2} \log^{3/2} N \left(\frac{1}{M} + \frac{N}{M^e}\right)^{\frac{1}{2e(e-1)+1}}.$$

Proof. This result is based on a theorem proved in [21] which we apply to evaluate $S(a, M)$. We give here a specific version adapted to our case:

Theorem 5. *Let e be an integer with $e \geq 2$, and let $a/N \in \mathbb{R}$. Suppose that, for some $c \in \mathbb{Z}$ and $N \in \mathbb{N}$ with $\gcd(c, N) = 1$, one has $|a/N - c/N| \leq N^{-2}$ and $N \leq M^e$. Then one has*

$$\sum_{1\leq x\leq M} e(ax^e/N) \ll M^{1+\epsilon}(N^{-1} + M^{-1} + N \cdot M^{-e})^{\sigma(e)},$$

where $\sigma(e)^{-1} = 2e(e-1)$.

According to [21], the factor M^ϵ may be replaced by $\log(2M)$, if one increases $\sigma(e)^{-1}$ from $2e(e-1)$ to $2e^2-2e+1$. For the sake of simplicity, we bound $\log(2M)$

by $\log N$ (with the weak assumption that $M \leq N/2$) and we neglect the term N^{-1} since it is negligible compared to $\min(M^{-1}, N \cdot M^{-e})$. Thus we obtain:

$$|S(a, M)| \ll M \log N (M^{-1} + N \cdot M^{-e})^{\frac{1}{2e(e-1)+1}},$$

and the replacement of B in Lemma 4 with this evaluation concludes the proof. □

Since Theorem 4 does not require any assumption of size on M, we want to define this parameter using the optimal bound $N^{1/e}$. In other words, we write M as $M = N^{1/e} \log^c N$ with $c > 0$. We propose in this case two corollaries depending on the value of c proved in appendices A.3 and A.4.

Corollary 1. *Let $N = pq$ the product of two large primes and M an integer such as $M = N^{1/e} \log^c N$ and $c \leq \frac{1}{e(e-1)} \frac{\log N}{\log \log N}$. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and U_k is:*

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2}}{(\log N)^{ce\sigma(e)-3/2}},$$

with $\sigma^{-1}(e) = 2e(e-1) + 1$.

Note that the statistical distance will be negligible if $ce\sigma(e) - 3/2 \gg 0$, meaning if $c \gg 3(e-1) + 3/2e$.

Corollary 2. *Let $N = pq$ the product of two large primes and M an integer such as $M = N^{1/e} \log^c N$ and $\frac{1}{e(e-1)} \frac{\log N}{\log \log N} \leq c < \frac{e-1}{e} \frac{\log N}{\log \log N}$. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and U_k is:*

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2}}{N^{\frac{\sigma(e)}{e}} (\log N)^{c\sigma(e)-3/2}},$$

with $\sigma^{-1}(e) = 2e(e-1) + 1$.

An interesting value of c is $c = (\frac{1}{2} - \frac{1}{e}) \frac{\log N}{\log \log N}$ which represents the case $M = \sqrt{N}$. For $e = 3$, this is the lower bound of c in Corollary 2 and for $e > 3$ it is included in the defined interval.

Let us give a numerical example for Corollary 1, the most interesting corollary since it treats values of M as close as possible to the optimal bound $N^{1/e}$. We consider classical cryptographic parameters for the upper bound of δ, i.e. 2^{-80}, and we put $e = 3$. Suppose that we want to have a negligible statistical distance for $k = 160$, then a modulus of 4096 bits leads to an impossibility. Indeed, Corollary 1 requires that $7 \ll c < 56$ and the result on δ is true for $c \geq 65$. However, with a modulus of 8192 bits one obtains the condition $60 \leq c < 105$. In other words, with an input of at least 3511 bits for the function f, the 160 least significant bits of $f(x)$ are statistically indistinguishable from the uniform distribution on $\{0, 1\}^{160}$.

To conclude, we extend ours theorems and corollaries to another but similar case proved in appendix A.5.

Corollary 3. *Let $N = pq$ the product of two large prime integers and M an integer less than N. The results from Theorem 3 and Theorem 4 on the statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and U_k are still valid for $\mathsf{msb}_k(x^e \bmod N)$.*

4 Applications of these bounds to cryptographic cases

4.1 Security of Micali-Schnorr pseudo-random number generator

Micali-Schnorr PRNG. A pseudorandom generator is a deterministic polynomial time algorithm that expands short seeds (made of truly random bits) into longer bit sequences, whose distribution cannot be distinguished from uniformly random bits by a computationally bounded algorithm.

Let (e, N) a RSA public key with e small compared to $\log N$ and $x_0 \in [0, 2^r)$ with $2^r \ll N$ a secret seed of size r. The Micali-Schnorr pseudorandom generator proposed in [18] is defined as follows:

$$v_i = x_{i-1}^e \mod N,$$
$$v_i = 2^k x_i + w_i \quad \text{for } i \geq 1.$$

At each iteration, this generator outputs the k least significant bits of v_i, denoted by w_i. In addition, denoting n the size of the modulus N, only x_i of size $r = n-k$, unknown, is reused for the next iteration. Since the generator outputs $O(k/\log e)$ bits per multiplication, one wants k to be as large as possible and e to be as small as possible.
This pseudorandom generator is proven secure under the following strong assumption:

Assumption 1 *The distribution of $x^e \bmod N$ for random r-bit integers x is indistinguishable by all polynomial-time statistical tests from the uniform distribution of elements of $\mathbb{Z}_N^*$.*

Clearly this assumption cannot be true if one does not restrain the tests to polynomial-time ones only because of the lack of entropy in input. Micali and Schnorr have proposed to use the parameters $r = 2n/e$ and thus $k = n(1 - 2/e)$ which are very aggressive parameters in order to raise the efficiency of the generator. However, they just assumed that using these parameters, it may be sufficient for the indistinguishability between the two sets we are focusing on.

Our result. We do not contradict this assumption since our theorems give upper bounds on δ but they can be used to define the sizes of parameters k and r such as the the statistical distance is bounded as desired. Note that we study a single iteration of the generator as in [12] for example, the consideration of two or more consecutive outputs being a difficult task. The proof of the following corollary is in appendix A.6

Corollary 4. *Let (e, N) a RSA public key with e small compared to $\log N$ and d a security parameter such that $\delta < 2^{-d}$. Let $\alpha \in (0,1)$ such that Micali-Schnorr pseudo-random number generator outputs the $(1-\alpha)\log N$ least significant bits at one iteration. This output is indistinguishable from the uniform distribution on $\{0,1\}^{(1-\alpha)\log N}$ if*

$$\alpha > 2/3 + \frac{2d + 4\log e}{3\log N} + \frac{\log\log N}{\log N}.$$

Note that Theorem 4 which treats in a relevant way the case $\alpha < 1/2$, is useless for this application because of the necessarily link between k and M: if $M = N^{\alpha}$ then $2^k \simeq N^{1-\alpha}$. Indeed, we start from the following general bound:

$$\delta \leq \sqrt{\frac{2^k}{N}} + 2^{k/2}\log^{3/2} N\left(\frac{1}{M} + \frac{N}{M^e}\right)^{\frac{1}{2e(e-1)+1}}.$$

By using the same notations for M and 2^k and bounding the right term by 2^{-d}, we first consider the case $N^{-\alpha} < N^{1-\alpha e}$ (i.e. $1/M < N/M^e$) and neglect the factor 2 (as in the proofs of the corollaries) and the term $\sqrt{\frac{2^k}{N}}$. Thus we obtain $\alpha > A(N,e)/B(e)$ with:

$$A(N,e) = \frac{3/2\log\log N}{\log N} + \frac{1}{2} + \frac{1}{2e(e-1)+1} + \frac{d}{\log N},$$

$$B(e) = \left(\frac{1}{2} + \frac{e}{2e(e-1)+1}\right).$$

For a fixed e, the function $A(N,e)/B(e)$ is decreasing in N and tends to:

$$\frac{2e(e-1)+3}{2e^2+1},$$

which is larger than $2/3$ whatever is the value of the public RSA exponent $e \geq 3$. The same result can be achieved by considering the case $N^{-\alpha} > N^{1-\alpha e}$.

It is interesting to note that when N tends to infinity, this bound tends to $2/3$. In other words we cannot expect to have a positive result of indistinguishability according to our results if one outputs more than $(\log N)/3$ of the least significant bits asymptotically. As concrete example, for $N = 2^{1024}, e = 3$ and $d = 80$, that gives an input greater than 747 bits (and thus an output lesser than 277 bits).

4.2 Semantic Security of PKCS #1 v1.5 encryption

RSA is a well-known asymmetric cryptosystem, first publicized in 1977. Even nowadays it is frequently used in applications where security of digital data is a concern. However the basic RSA encryption process, meaning without any padding of the plaintext, is vulnerable to quite simple or clever attacks (see for example [10,15]). The standard PKCS #1 v1.5 proposes a padding which parries a part of these attacks, nevertheless it has been defeated by Bleichenbacher in [6].

PKCS #1 v1.5 encryption. We recall the encryption scheme proposed in the standard PKCS #1 v1.5. By denoting (e, N) the RSA public key with n the size of the modulus N, a message m of size at most $\ell = n - 88$ is padded as follows:

$$m_{padded} = 00||02||PS||00||m$$

with PS the padding string of size $r = n - \ell - 24$. Then, m_{padded} is encrypted using the RSA encryption process.

Semantic Security. The semantic security requires that the adversary should not gain any advantage or information from having seen the cipher text resulting from an encryption algorithm. This can be formalized by this concrete definition:

Definition 3. *An encryption scheme (Enc, Dec) is (t, o, ϵ) semantically secure if for every distribution X over messages, every functions $I : \{0,1\}^m \to \{0,1\}^*$ and $f : \{0,1\}^m \to \{0,1\}^*$ (of arbitrary complexity) and every function A of complexity $t_A \leq t$, there is a function A' of complexity $\leq t_a + o$ such that*

$$\left| \Pr[A(Enc(K, M), I(m)) = f(M)] - \Pr[A'(I(m)) = f(M)] \right| \leq \epsilon.$$

$I(m)$ can be seen as the knowledge of the adversary on the message M, whereas $f(M)$ represents the knowledge the adversary would learn.

Our result. We are interested by bounding the statistical distance between the function $\mathsf{lsb}_k(Pad(x)^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_{2^r}$ and the uniform distribution on $\{0,1\}^k$, x being an integer whose binary representation is PS. More precisely we define $Pad(x)$ as $Pad(x) = 2^{n-16}b + 2^{l+8}x + m$ with b of size 8 and m of size ℓ two integer fixed. This application can be viewed as a partial study of the semantic security of PKCS #1 v1.5 encryption because we only give access to the k least significant bits of the ciphertext.

Corollary 5. *Let $N = pq$ the product of two large prime integers and M an integer less than N. Let $Pad(x)$ a function defined same as above. The results from Theorem 3 and Theorem 4 on the statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and the uniform distribution on $\{0,1\}^k$ are still valid for $\mathsf{lsb}_k(Pad(x)^e \bmod N)$.*

With a security parameter $d = 80$, $\log N = 1024$, $e = 3$ and $\log M = 872$ (we consider the encryption of a symmetric cryptosystem key of size 128), we obtain the condition $k < 524$.

References

1. Gilles Barthe, Benjamin Grégoire, Yassine Lakhnech, and Santiago Zanella Béguelin. Beyond provable security verifiable ind-cca security of oaep. In Aggelos Kiayias, editor, *CT-RSA*, volume 6558 of *Lecture Notes in Computer Science*, pages 180–196. Springer, 2011.

2. Aurélie Bauer, Jean-Sébastien Coron, David Naccache, Mehdi Tibouchi, and Damien Vergnaud. On the Broadcast and Validity-Checking Security of pkcs#1 v1.5 Encryption. In Jianying Zhou and Moti Yung, editors, *ACNS*, volume 6123 of *Lecture Notes in Computer Science*, pages 1–18, 2010.
3. Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In Dorothy E. Denning, Raymond Pyle, Ravi Ganesan, Ravi S. Sandhu, and Victoria Ashby, editors, *ACM Conference on Computer and Communications Security*, pages 62–73. ACM, 1993.
4. Mihir Bellare and Phillip Rogaway. Optimal asymmetric encryption. In Alfredo De Santis, editor, *EUROCRYPT*, volume 950 of *Lecture Notes in Computer Science*, pages 92–111. Springer, 1994.
5. Mihir Bellare and Phillip Rogaway. The exact security of digital signatures - how to sign with rsa and rabin. In Ueli M. Maurer, editor, *EUROCRYPT*, volume 1070 of *Lecture Notes in Computer Science*, pages 399–416. Springer, 1996.
6. Daniel Bleichenbacher. Chosen ciphertext attacks against protocols based on the rsa encryption standard pkcs #1. In *CRYPTO*, pages 1–12, 1998.
7. Lenore Blum, Manuel Blum, and Mike Shub. A simple unpredictable pseudo-random number generator. *SIAM J. Comput.*, 15(2):364–383, 1986.
8. Dan Boneh and Ramarathnam Venkatesan. Breaking rsa may not be equivalent to factoring. In Kaisa Nyberg, editor, *EUROCRYPT*, volume 1403 of *Lecture Notes in Computer Science*, pages 59–71. Springer, 1998.
9. Don Coppersmith. Small solutions to polynomial equations, and low exponent rsa vulnerabilities. *J. Cryptology*, 10(4):233–260, 1997.
10. Don Coppersmith, Matthew K. Franklin, Jacques Patarin, and Michael K. Reiter. Low-exponent rsa with related messages. In *EUROCRYPT*, pages 1–9, 1996.
11. Pierre-Alain Fouque, Damien Vergnaud, and Jean-Christophe Zapalowicz. Time/memory/data tradeoffs for variants of the rsa problem. In Ding-Zhu Du and Guochuan Zhang, editors, *COCOON*, volume 7936 of *Lecture Notes in Computer Science*, pages 651–662. Springer, 2013.
12. John Friedlander and Igor Shparlinski. On the distribution of the power generator. *Math. Comput.*, 70(236):1575–1589, 2001.
13. Eiichiro Fujisaki, Tatsuaki Okamoto, David Pointcheval, and Jacques Stern. Rsa-oaep is secure under the rsa assumption. In Kilian [16], pages 260–274.
14. Rosario Gennaro. An improved pseudo-random generator based on discrete log. In *CRYPTO*, pages 469–481, 2000.
15. Johan Håstad. Solving simultaneous modular equations of low degree. *SIAM J. Comput.*, 17(2):336–341, 1988.
16. Joe Kilian, editor. *Advances in Cryptology - CRYPTO 2001, 21st Annual International Cryptology Conference, Santa Barbara, California, USA, August 19-23, 2001, Proceedings*, volume 2139 of *Lecture Notes in Computer Science*. Springer, 2001.
17. Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Cambridge University Press, 1996.
18. Silvio Micali and Claus-Peter Schnorr. Efficient, perfect polynomial random number generators. *J. Cryptology*, 3(3):157–172, 1991.
19. Victor Shoup. Oaep reconsidered. In Kilian [16], pages 239–259.
20. Victor Shoup. *A computational introduction to number theory and algebra*. Cambridge University Press, 2006.
21. TD Wooley. Vinogradov's mean value theorem via efficient congruencing. *Annals of Mathematics*, 175(3):1575–1627, 2012.

A Proofs omitted

A.1 Proof of Lemma 3

Lemma 3. *Let $N = pq$ a RSA modulus with p, q two large prime numbers and let u, v be such that $uq = 1 \bmod p$ and $vp = 1 \bmod q$. Then, for any polynomial $f \in \mathbb{Z}[X]$ with integer coefficients:*

$$\sum_{j=0}^{N-1} e_N(f(j)) = \sum_{j_1=0}^{p-1} e_p(uf(j_1)) \sum_{j_2=0}^{q-1} e_q(vf(j_2)).$$

Proof. Starting from the modular equality $uq + vp = 1$, one has $j = juq + jvp \bmod N$ with $0 \leq j \leq N-1$ and more precisely:

$$j = j_1 uq + j_2 vp \bmod N,$$

with $0 \leq j_1 \leq p-1$ and $0 \leq j_2 \leq q-1$. The binomial theorem implies:

$$\begin{aligned} j^k &= \sum_{i=0}^{k} \binom{k}{i} (j_1 uq)^{k-i} (j_2 vp)^i \bmod N, \\ &= (j_1 uq)^k + (j_2 vp)^k \bmod N, \\ &= uqj_1^k + vpj_2^k \bmod N, \end{aligned}$$

using the fact that all the monomials contain a factor pq except for the first and the last one in the first equality. Moreover, for $k \geq 1, (uq)^k = uq \bmod N$ in the second equality since $(uq)^2 = uq(1 - vp) = uq \bmod N$ and by induction on k. Then, by denoting $f(j) = \sum_{k=0}^{d} a_k j^k$ with $a_k \in \mathbb{Z}$ we obtain $f(j) = uqf(j_1) + vpf(j_2)$ with $0 \leq j \leq N-1$, $0 \leq j_1 \leq p-1$, $0 \leq j_2 \leq q-1$ and thus:

$$\sum_{j=0}^{N-1} e(2i\pi f(j)/N) = \sum_{j_1=0}^{p-1} e(2i\pi uqf(j_1)/N) \sum_{j_2=0}^{q-1} e(2i\pi vpf(j_2)/N).$$

□

A.2 Proof of Lemma 5

Lemma 5. *If for all $a, b \in \mathbb{Z}^2$ and for any polynomial $P \in \mathbb{Z}[X]$ with integer coefficients,*

$$\left| \sum_{x \in \mathbb{Z}_N} e_N(aP(x)) e_N(bx) \right| \leq C$$

then, for $0 \leq M \leq N$,

$$\left| \sum_{1 \leq x \leq M} e_N(aP(x)) \right| \leq C \log N.$$

Proof. Indeed, we have the following equality:

$$\left|\sum_{1\leq x\leq M} e_N(aP(x))\right| = \left|\sum_{x\in\mathbb{Z}_N} \mathbb{1}_{[1,M]} e_N(aP(x))\right|$$

with $\mathbb{1}_{[1,M]} = 1/N \sum_{b=1}^{N}\sum_{m=1}^{M} e_N(b(x-m))$. Thus, we obtain:

$$\begin{aligned}\left|\sum_{1\leq x\leq M} e_N(aP(x))\right| &= \left|1/N \sum_{b=1}^{N}\sum_{x\in\mathbb{Z}_N} e_N(aP(x)) e_N(bx) \sum_{m=1}^{M} e_N(-bm)\right| \\ &\leq C/N \sum_{b=1}^{N}\left|\sum_{m=1}^{M} e_N(-bm)\right| \\ &\leq C/N \times N\log N \leq C\log N.\end{aligned}$$

A.3 Proof of Corollary 1

Corollary 1. *Let $N = pq$ the product of two large primes and M an integer such as $M = N^{1/e}\log^c N$ and $c \leq \frac{1}{e(e-1)}\frac{\log N}{\log\log N}$. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and the uniform distribution on $\{0,1\}^k$ is:*

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2}}{(\log N)^{ce\sigma(e)-3/2}},$$

with $\sigma^{-1}(e) = 2e(e-1)+1$.

Proof. In order to simplify the formula, we consider here that $\max(1/M, N/M^e) = N/M^e$ and we write $(1/M + N/M^e)$ as N/M^e (we will consider that the factor 2 is included in the majoration of $\log 2M$). Thus we obtain the following bound on δ:

$$\delta \leq \sqrt{\frac{2^k}{N}} + 2^{k/2}\log^{3/2} N\left(\frac{N}{M^e}\right)^{\frac{1}{2e(e-1)+1}}$$

and one replaces M by $M = N^{1/e}\log^c N$ to get the result. Finally, because $1/M \leq N/M^e$ then we obtain a condition on c:

$$N^{\frac{e-1}{e}}\log^{c(e-1)} N \leq N,$$

which is equivalent to

$$c \leq \frac{1}{e(e-1)}\frac{\log N}{\log\log N}.$$

□

A.4 Proof of Corollary 2

Corollary 2. *Let $N = pq$ the product of two large primes and M an integer such as $M = N^{1/e}\log^c N$ and $\frac{1}{e(e-1)}\frac{\log N}{\log\log N} \leq c < \frac{e-1}{e}\frac{\log N}{\log\log N}$. The statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and the uniform distribution on $\{0,1\}^k$ is:*

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2}}{N^{\frac{\sigma(e)}{e}}(\log N)^{c\sigma(e)-3/2}},$$

with $\sigma^{-1}(e) = 2e(e-1)+1$.

Proof. Contrary to the first corollary, we consider here that $N/M^e \leq 1/M$ and we apply the same simplification. We thus obtain a lower bound on c:

$$c \geq \frac{1}{e(e-1)}\frac{\log N}{\log\log N},$$

and the condition $M < N$ implies an upper bound:

$$c < \frac{e-1}{e}\frac{\log N}{\log\log N}.$$

The bound of δ becomes:

$$\delta \leq \sqrt{\frac{2^k}{N}} + 2^{k/2}\log^{3/2} N\left(\frac{1}{M}\right)^{\frac{1}{2e(e-1)+1}}$$

Finally, one replaces M by $M = N^{1/e}\log^c N$ and gets the result.

□

A.5 Proof of Corollary 3

Corollary 3. *Let $N = pq$ the product of two large prime integers and M an integer less than N. The results from Theorem 3 and Theorem 4 on the statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and the uniform distribution on $\{0,1\}^k$ are still valid for $\mathsf{msb}_k(x^e \bmod N)$.*

Proof. Indeed, the single difference appears in the evaluation of $\mathrm{Col}(X_{M,e,N,k})$:

$$\begin{aligned}\mathrm{Col}(X_{M,e,N,k}) &= \frac{1}{M^2} \times \left|\{(x,y) \in [0, M-1]^2 \mid \exists u \leq K, x^e - y^e = u \bmod N\}\right|,\\ &= \frac{1}{M^2 N}\sum_{x=0}^{M-1}\sum_{y=0}^{M-1}\sum_{u=0}^{K}\sum_{a=0}^{N-1} e_N(a(x^e - y^e - u)),\end{aligned}$$

with K still equal to $\left\lfloor \frac{N-1}{2^k} \right\rfloor$. Since

$$\sum_{a=1}^{N-1}\left|\sum_{u=0}^{K} e_N(-a2^k \cdot u)\right| = \sum_{a=1}^{N-1}\left|\sum_{u=0}^{K} e_N(-a \cdot u)\right|,$$

the rest of the proof remains the same. □

A.6 Proof of Corollary 4

Corollary 4. *Let (e, N) a RSA public key with e small compared to $\log N$ and d a security parameter such that $\delta < 2^{-d}$. Let $\alpha \in (0, 1)$ such that Micali-Schnorr pseudo-random number generator outputs the $(1-\alpha)\log N$ least significant bits at one iteration. This output is indistinguishable from the uniform distribution on $\{0,1\}^{(1-\alpha)\log N}$ if*

$$\alpha > 2/3 + \frac{2d + 4\log e}{3\log N} + \frac{\log\log N}{\log N}.$$

Proof. By denoting $M = N^{\alpha}$ with $\alpha < 1$ and $2^k \simeq N^{1-\alpha}$, we obtain from Theorem 3:

$$\delta \leq N^{-\alpha/2} + N^{1-\frac{3}{2}\alpha} e^2 \sqrt{N} \log^{3/2} N.$$

Since $-\alpha/2 < 1 - \frac{3}{2}\alpha$ is equivalent to $\alpha < 1$ and $e^2\sqrt{N}\log^{3/2} N$ is quite large for cryptographic parameters, we will neglect the first term. We bound the statistical distance by 2^{-d} with d a security parameter and obtain the following condition on α:

$$\alpha > 2/3 + \frac{2d + 4\log e}{3\log N} + \frac{\log\log N}{\log N}.$$

□

A.7 Proof of Corollary 5

Corollary 5. *Let $N = pq$ the product of two large prime integers and M an integer less than N. Let $Pad(x)$ a function defined same as above. The results from Theorem 3 and Theorem 4 on the statistical distance between $\mathsf{lsb}_k(x^e \bmod N)$ for x randomly chosen in $\mathbb{Z}_M$ and the uniform distribution on $\{0,1\}^k$ are still valid for $\mathsf{lsb}_k(Pad(x)^e \bmod N)$.*

Proof. The result of Lemma 4 remains valid for this application. Indeed, the characteristic function becomes:

$$\mathbb{1}(x, y, u) = 1/N \times \sum_{a=0}^{N-1} e_N(a(Pad(x)^e - Pad(y)^e - 2^k \cdot u)),$$

and by denoting $M = 2^r$, $S(a, M)$ is now defined as:

$$S(a, M) = \sum_{0 \leq x < M} e_N(aPad(x)^e).$$

We still denote by B the bound on $S(a, M)$ for any integer a and get:

$$\delta \leq \sqrt{\frac{2^k}{N}} + \frac{2^{k/2} B \log^{1/2} N}{M}$$

The analyze of B is quite simple since Lemma 3 implies:

$$\left| \sum_{x \in \mathbb{Z}_N} e_N(aPad(x)^e) \right| = \left| \sum_{x \in \mathbb{F}_p} e_p(v \times aPad(x)^e) \sum_{x \in \mathbb{F}_q} e_q(u \times aPad(x)^e) \right|.$$

with $u, v \in \mathbb{Z}$ such as $up + vq = 1$ and the Weil bound gives the same result thanks to the degree of the polynomial $Pad(x)^e$ equal to e:

$$\left| \sum_{x \in \mathbb{Z}_N} e_N(aPad(x)^e) \right| \leq ep^{1/2} \times eq^{1/2} \leq e^2 N^{1/2}.$$

Thus Theorem 3 is still valid for this application. Finally Theorem 4 is correct with the polynomial $Pad(x)^e$ too (see its general form in [21]). $\square$