Regularity of Lossy RSA on Subdomains and Its Applications

Mark Lewko[1], Adam O'Neill[2], and Adam Smith[3]

[1] University of California, Los Angeles
mlewko@math.ucla.edu
[2] Boston University
amoneill@bu.edu
[3] Pennsylvania State University
asmith@cse.psu.edu

Abstract. We build on an approach of Kiltz et al. (CRYPTO '10) and bring new techniques to bear on the study of how "lossiness" of the RSA trapdoor permutation under the Φ-Hiding Assumption (ΦA) can be used to understand the security of classical RSA-based cryptographic systems. In particular, we show that, under ΦA, several questions or conjectures about the security of such systems can be reduced to bounds on the regularity (the distribution of the primitive e-th roots of unity mod N) of the "lossy" RSA map (where e divides $\phi(N)$). Specifically, this is the case for: (i) showing that large consecutive runs of the RSA input bits are simultaneously hardcore, (ii) showing the widely-deployed PKCS #1 v1.5 encryption is semantically secure, (iii) improving the security bounds of Kiltz et al. for RSA-OAEP. We prove several results on the regularity of the lossy RSA map using both classical techniques and recent estimates on Gauss sums over finite subgroups, thereby obtaining new results in the above applications. Our results deepen the connection between "combinatorial" properties of exponentiation in $\mathbb{Z}_N$ and the security of RSA-based constructions.

Keywords: RSA encryption, PKCS #1 v1.5, Lossy trapdoor functions, Φ-Hiding Assumption, Gauss sums.

1 Introduction

Cryptographic systems built from the RSA trapdoor permutation [34] are ubiquitous in practice. Though these schemes are simple and highly efficient, they are typically only proven secure in the random oracle model [3], if at all.[1] An important research direction is to prove their security under better-understood assumptions. For example, consider the "simple embedding" RSA-based encryption scheme specified by RSA PKCS #1 v1.5, which is still in widespread use: roughly, the encryption of a plaintext x is $f_{N,e}(x, r) = (x\|r)^e \bmod N$, where r

[1] There are more recent constructions without random oracles, e.g., [19,20], but they are less efficient and seem unlikely to be used in practice in the near future.

T. Johansson and P. Nguyen (Eds.): EUROCRYPT 2013, LNCS 7881, pp. 55–75, 2013.

is a random string of appropriate length and '$\|$' denotes string concatenation.[2] There was until now no proof of security of this scheme under a standard and well-studied assumption on RSA.[3] In this paper, we show that the security of this and related constructions can be analyzed under natural assumptions without the need for the random oracle model. Our analysis relies on new connections between the security of RSA and "combinatorial" properties of arithmetic in $\mathbb{Z}_N$ (namely, the regularity of exponentation on arithmetic sequences and bounds on the magnitude of Gauss sums).

KEY TOOL: LOSSINESS. A keyed trapdoor function family f_{pk} on k-bits is *L-lossy* [32] if there are two algorithms, or "modes", for generating public keys: one which generates key *pairs* (pk, sk) such that f_{pk} is injective and can be inverted efficiently given sk, and one which generates keys pk for which f_{pk} is "L-lossy", meaning the image of f_{pk} has at most 2^{k-L} points. The security requirement is that the two modes be computationally indistinguishable. The concept has found applications in many areas of cryptography, for example in [32,6,1,2,31].

Lossiness is a powerful tool, since it allows one to prove security with respect to the lossy mode, where information-theoretic techniques often apply. As concrete example, one can show that a lossy function family admits many simultaneously hard-core bits: in the lossy mode, the (average) min-entropy of a uniformly random input X given $f_{pk}(X)$ is at least L, and hence one can use an appropriate randomness extractor Ext, such as a 2-wise-independent hash family, to obtain a $L - 2\log(1/\varepsilon)$-bit string that is ε-close to uniform even given $f_{pk}(X)$ and the seed. In the injective mode, this string will be *pseudorandom* given $f_{pk}(X)$ and the seed, since a distinguisher for the extracted string would imply a distinguisher that tells apart lossy/injective keys. The existence of many simultaneously hardcore bits allows for the design of efficient, semantically-secure encryption schemes (say, by using these bits as a one-time pad).

USING LOSSINESS TO ANALYZE CLASSICAL RSA-BASED CONSTRUCTIONS. Lossiness has mostly been used in the literature as a tool for designing new cryptographic systems. Recently, however, Kiltz et al. [25] showed that the concept also sheds light on existing, widely-used constructions. Specifically, they showed that RSA-OAEP [4] is semantically secure under the ϕ-Hiding Assumption. The ϕ-Hiding Assumption, abbreviated ΦA, states (roughly) that given an RSA modulus $N = pq$, it is hard to distinguish primes that divide $\phi(N) = (p-1)(q-1)$ from those that do not. ΦA has been used as the basis for a number of efficient protocols [10,9,14,18]. It has also attracted attention of cryptanalysis: the current best attack uses Coppersmith's techniques [12] and applies when $e \leq p^{1/2-\varepsilon}$; other attacks [35] are for moduli of a special form that does not include RSA. Kiltz et al. [25] observed that the RSA map $x \mapsto x^e$ in $\mathbb{Z}_N^*$ is $\log(e)$-*lossy* under

[2] In practice x and r are typically switched and some bits of r are a fixed constant; however, this won't affect our results.

[3] We clarify that the parameters (i.e., the RSA modulus and exponent length) supported by our security proof are not practical (see the discussion at the end of the Intro for more details). However, prior work does not provide a proof for *any* parameter settings.

ΦA: Valid RSA keys (N, e), for which $\gcd(e, \phi(N)) = 1$, are computationally indistinguishable from "lossy" pairs (N, e) for which e divides $p - 1$, and the map $x \mapsto x^e$ in $\mathbb{Z}_N^*$ is e-to-one when e divides $(p - 1)$.

Thus, ΦA implies that the RSA map hides $\log(e)$ bits of information (on average) about x. But without additional information about *which* bits are hidden, it seems we can only analyze constructions which have extractor-like objects (such as keyed hash functions), explicitly built in. For example, Kiltz et al. [25] analyzed the OAEP padding scheme by modeling the random oracles as keyed, t-wise independent functions. One can similarly exploit the observation above about simultaneous hardcore bits to replace the random oracle in the "Simple RSA" or "RSA-KEM" scheme proposed by Shoup [36]. However, this methodology does not apply to schemes that do not use keyed hash functions or do not use hash functions at all.

1.1 Our Contributions

We show specific, natural functions that are hidden by RSA in the lossy case, and use these results to obtain proofs of several natural conjectured security properties of RSA-based constructions, assuming ΦA. Roughly, the three main applications are: (i) Any run of $\log(e) = \Omega(\log N)$ consecutive physical bits are *simultaneously* hardcore for RSA. (The assumption that RSA is hard to invert implies only that $\log \log N$ physical bits are simultaneously hardcore.) (ii) PKCS #1 v1.5 (described above) is semantically secure against chosen *plaintext* attacks (CPA). (iii) Improved parameters for the CPA security of RSA-OAEP (improving on the reduction in [25]).

These results emanate from our core technical contribution: showing that, in the lossy setting (when e divides $\phi(N)$), exponentiation by e is nearly *regular* on certain subdomains $\mathcal{K} \subseteq \mathbb{Z}_N$. Regularity means that all (or most) points in the image of $x \mapsto x^e$ have approximately the same number of preimages in $\mathcal{K}$. This implies in turn that X^e is approximately uniform on its image when X is uniform on $\mathcal{K}$. Consider, for example, the natural conjecture that the t most significant bits of the input are hardcore for exponentiation. To prove this conjecture under ΦA, it suffices to show that, for every fixed string $z \in \{0,1\}^t$, the value $(z \| R)^e$ is nearly uniform, where $R \leftarrow \{0,1\}^{\lceil \log N \rceil - t}$ (this implies that any two settings z, z' of the hardcore bits are statistically indistinguishable in the lossy mode, and computationally indistinguishable in the usual injective mode). This question corresponds to the regularity of exponentiation by e on the arithmetic progression $\mathcal{K} = \{z 2^{\lceil \log N \rceil - t} + r : r = 0, ..., 2^{\lceil \log N \rceil - t} - 1\}$.

Below, we explain our results in more detail.

Results on Regularity of Exponentiation. We prove several results on the regularity of lossy exponentiation on subdomains of $\mathbb{Z}_N$ when $N = pq$. The subdomains we consider have some additive structure, which somehow breaks up the multiplicative structure of exponentiation. Consider a subdomain $\mathcal{K} \subseteq \mathbb{Z}_N$. Let X be uniform on $\mathcal{K}$ and U be uniform on $\mathbb{Z}_N$. Note that for X^e to be close to U^e, the set $\mathcal{K}$ must have size at least $\phi(N)/e \approx N/e$.

Regularity for Random Translations of a Set: Our first result considers random translations of an arbitrary subdomain $\mathcal{K}$. Specifically, we show that the pair $(C, (C+X)^e)$ is statistically close to (C, U^e) when C is uniform on $\mathbb{Z}_n$, as long as $\mathcal{K}$ has size larger than N/e. The proof relies on a careful collision probability argument. One key piece of that argument is the observation that the random offset and exponentiation *together* behave like a universal hash function: for any two values $a, b \in \mathbb{Z}_N$, the ratio $\left(\frac{a+C}{b+C}\right)^e$ is nearly uniformly distributed over e-th residues, conditioned on the denominator being invertible. The other main piece is a careful accounting of the probability of noninvertible elements; this is delicate because the conversion from collision probability to $\mathcal{L}_1$-distance can amplify very small irregularities.

This first result achieves optimal parameters but it takes advantage of "averaging" in two senses: first, it averages over translations of the initial set $\mathcal{K}$ and second, it only implies regularity on average over points in the pre-image (since we show that the $\mathcal{L}_1$-distance between the resulting distributions is small). This turns out to be insufficient for some applications, motivating our second result.

Regularity for Arithmetic Progressions and the Relation to Gauss Sums: Our second result is more specific: we show that if $\mathcal{K}$ is a sufficiently long arithmetic progression (with period relatively prime to N), then exponentiation is regular on $\mathcal{K}$ in a strong sense: the number of preimages of *every* point in the image is approximately the same.

The main tool in our analysis is a reduction from the question of regularity to bounds on Gauss sums. Given a prime p, and integers a, d, consider the sum $\mathcal{G}_p(a,d) := \sum_{x=1}^{p} \omega^{ax^d}$, where $\omega = \exp(2\pi i/p)$ is a primitive p-th root of unity and the arithmetic is in $\mathbb{C}$. We show that if e divides $p-1$ and $N = pq$, then

$$\left|\Pr(X^e = a) - \frac{e}{N}\right| \leq \max_{b \neq 0} \left|\mathcal{G}_p(b, \tfrac{p-1}{e})\right| \cdot O\left(\frac{e \log K}{pK}\right)$$

where K is the length of $\mathcal{K}$. One can think of the Gauss sums $\mathcal{G}_p(\cdot, d)$ as the Fourier coefficients of the function $x \mapsto x^d$ over $\mathbb{Z}_p$. The proof of our main lemma uses Fourier analysis over $Z_N = \mathbb{Z}_p \times \mathbb{Z}_q$ to connect regularity to the magnitude of the sum.

Leveraging the rich literature on bounds on Gauss sums, we obtain regularity results for different ranges of e (relative to p). These results show that exponentiation (when e divides $\phi(N)$) is a *deterministic extractor* for sources whose support is a sufficiently long arithmetic progression. Moreover, the output distributions are close to uniform not only in $\mathcal{L}_1$-distance but also in the stronger $\mathcal{L}_\infty$ sense. Given the state of our knowledge of bounds for Gauss sums, this second class of results yields weaker (but still useful) bounds on uniformity than our first result. (For a comparison of the bounds, see the end of Section 4.)

Applications to RSA-Based Cryptosystems. Our regularity bounds imply the following new security results for RSA-based constructions under the ϕ-Hiding Assumption (ΦA):

Natural hardcore bits for RSA: Any run of about $\log(e)$ consecutive physical bits of x are simultaneously hardcore for RSA. Specifically, let $f(x)$ denote a run of $\log(e) - 4\log(1/\varepsilon)$ physical bits of the input x. Our result on random translations implies (with some additional work) that the pair $(f(C), C^e)$ is ε-close to $(f(C), U^e)$ when e divides $p-1$ and C, U are uniform in $\mathbb{Z}_n$. If we consider either the most significant bits or least significant bits of x, then one can improve the run length to $\log(e) - 3\log(1/\varepsilon)$ (that is, we get $\log(1/\varepsilon)$ additional hardcore bits).

Semantic security of PKCS #1 v1.5: Our bounds on regularity on arithmetic progressions imply that PKCS #1 v1.5, which encrypts m as $(0002_{16} \| m \| 00_{16} \| R)^e$ where R is uniform, is semantically secure (aka. chosen-plaintext secure or IND-CPA) [16] for certain parameters. Indeed, note that the space of pre-images for a given message m forms an arithmetic progression with period 1 (variants of this scheme mentioned in Footnote 2 give progression with period 2^t for $t > 1$, which is still relatively prime to N). Under ΦA, with appropriately long R, the ciphertext is thus indistinguishable from uniform, for every fixed message m.

Improved Security for RSA-OAEP: Finally, our regularity bounds for arithmetic sequences also give tighter standard-model security proofs for the IND-CPA security of RSA-OAEP, improving on the results of [25]. The RSA-OAEP scheme as per PKCS #1 v2.1 encrypts m as $(0002_{16} \| \mathsf{OAEP}(m))^e$ where OAEP is some randomized, invertible transformation. Recall [25] were able to obtain their best security bound in the case that the lossy RSA map is (close to) regular on the subdomain $\{0002_{16} \| x \mid x \in \{0,1\}^{k-16}\}$ and left it as an open problem to prove this. As this subdomain forms an arithmetic progression, we resolve this positively.

Discussion and concrete parameters. As mentioned above, our result on regularity for random translations averages in two senses. The first sense is sufficient to obtain results on hardcore bits, but not security of PKCS # 1 v1.5 encryption or RSA-OAEP. Roughly, this is because the elements in the subdomains contain fixed messages and constants as substrings. Getting better regularity bounds without additional randomness (i.e., eliminating the first sense) remains an interesting open problem. This is primarily a concern for the PKCS #1 v1.5 application, since we need regularity on arithmetic progressions of length $2^\rho = 2^{\Omega(k)}$ where ρ is the length of the random padding. In the RSA-OAEP application we use length $2^{k-16} = 2^{k+O(1)}$, for which our regularity bounds on arithmetic progressions give much better parameters.

To give an idea of the concrete parameters we obtain, for modulus length $k = 2048$ we get about 190 natural hardcore bits. In our security bound for PKCS #1 v1.5, for modulus length $k = 8192$ we support about 128-bit messages. Finally, in the RSA-OAEP application we get significant savings: e.g., secure encryption of about 100-bit longer messages than supported by [25] for modulus length $k = 2048$ (274-bit messages for 80-bit security rather than 160-bits). In terms of practice, we view our results mainly as providing a qualitative, theoretical

backing for in-use schemes at some parameter settings. We hope our techniques will prove useful in future work and the bounds will be improved.

1.2 Related Work

Following [25], Kakvi and Kiltz [23] showed that lossiness of RSA under ΦA is also useful to understand security of a classical RSA-based *signature*, giving improved security bounds for the RSA Full-Domain Hash signature scheme [3]. Jager et al. [21] recently provided a standard-model analysis of TLS-DHE, another widely used protocol. Gauss sums also have applications to elliptic-curve cryptography, see e.g. [37,26,38].

Bleichenbacher [5] (see also [22]) gave a well-known chosen-ciphertext attack against PKCS #1 v1.5 encryption, which has since been patched and the scheme is still in widespread use for legacy reasons. Coron et al. [13] gave *chosen-plaintext* attacks on PKCS #1 v1.5 encryption. These do not contradict our results because the attacks of [13] are for different parameter settings. Specifically, they rely on the length of the random padding being quite small. Our results require sufficiently large random padding– at least $\frac{3}{4}\log N$ bits – as well as large e. Interestingly, our analysis implies a plausible setting in which PKCS #1 v1.5 is provably immune to the attacks of [13] under ΦA.

The "large hardcore bit conjecture" for RSA and the security of the simple embedding scheme are mentioned as important open problems by Goldreich [15]. Prior progress was made by Steinfeld et al. [39], who showed that the $1/2 - 1/e - \varepsilon - o(1)$ least significant bits of RSA are simultaneously hardcore under a computational problem related to the work of Coppersmith [12]. This result does not apply as such to PKCS #1 v1.5 because the latter does not use the full RSA domain (some bits are fixed constants). Moreover, we show *chosen-plaintext security*, i.e., security for arbitrary messages, rather than only for random ones (which, disregarding some of the other bits being fixed constants, is equivalent to the message bits being hardcore). The fact that PKCS #1 v1.5 is believed to be CPA-secure but no proof is known is also discussed by Katz and Lindell [24, pg. 363].

2 Preliminaries

Notation. For a probabilistic algorithm A, by $y \leftarrow\!\!\$\ A(x)$ we mean that A is executed on input x and the output is assigned to y, whereas if S is a finite set then by $s \leftarrow\!\!\$\ S$ we mean that s is assigned a uniformly random element of S. Unless otherwise specified, an algorithm may be probabilistic and its running-time includes that of any overlying experiment. We denote by 1^k the unary encoding of the security parameter k. We sometimes surpress dependence on k for readability. For $i \in \mathbb{N}$ we denote by $\{0,1\}^i$ the set of all (binary) strings of length i. If s is a string then $|s|$ denotes its length in bits, whereas if S is a set then $|S|$ denotes its cardinality. By '$\|$' we denote string concatenation. If s is a string then for all $1 \le i \le j \le |s|$ we denote by $s[i \ldots j]$ its substring of bits i through j.

We will occasionally use the usual asymptotic notation $X = O(Y)$ or $X \ll Y$ to indicate that $|X| \leq C|Y|$ for some unspecified constant C. We will also make use of the exact asymptotic notation $X = \mathcal{O}(Y)$ to indicate that $|X| \leq |Y|$ without any unspecified constant. We will use $\log(\cdot)$ to denote the natural logarithm and C to denote an absolute constant, which may change at unrelated occurrences. In addition, we use C_ε to denote an absolute constant that depends only on its subscript, ε. If X and Y are random variables on common domain, then their *statistical distance* is $\Delta(X,Y) = 1/2 \sum_x |\Pr[X = x] - \Pr[Y = x]| = 1/2 \cdot ||X - Y||_1$.

We borrow the following notation from [25]. For $i \in \mathbb{N}$ we denote by $\mathcal{P}_i$ the set of all i-bit primes. By RSA_k we denote the set of tuples (N,p,q) where p,q are distinct $k/2$-bit primes and $N = pq$. Let R be a relation on p and q. By $\mathsf{RSA}_k[R]$ we denote the subset of RSA_k for which the relation R holds on p and q. For example, let e be a prime. Then $\mathsf{RSA}_k[p = 1 \bmod e]$ is the set of all (N,p,q), where where $N = pq$ is the product of two distinct $k/2$-bit primes p,q and $p = 1 \bmod e$. That is, the relation $R(p,q)$ is true if $p = 1 \bmod e$ and q is arbitrary. By $(N,p,q) \leftarrow\!\!\$\, \mathsf{RSA}_k[R]$ we mean that (N,p,q) is sampled according to the uniform distribution on $\mathsf{RSA}_k[R]$.

RSA, Lossy RSA, and Phi-hiding. Recall that the *RSA function* for modulus N and encryption exponent e is defined as

$$\mathsf{RSA}_{N,e}(x) = x^e \bmod N\ .$$

To specify the RSA trapdoor permutation family we need to give a *parameter-generation algorithm* that specifies how parameters N,e are generated. (We ignore the decryption exponent d for now.) Letting $0 < c < 1$ be a public constant, we define two of them ("Injective RSA" and "Lossy RSA"):

Algorithm $\mathsf{RSA}_{inj}(1^k)$:	**Algorithm** $\mathsf{RSA}_{loss}(1^k)$:
$e \leftarrow\!\!\$\, \mathcal{P}_{ck}$	$e' \leftarrow\!\!\$\, \mathcal{P}_{ck}$
$(N,p,q) \leftarrow\!\!\$\, \mathsf{RSA}_k$	$(N',p',q') \leftarrow\!\!\$\, \mathsf{RSA}_k[p' = 1 \bmod e']$
Return (N,e)	Return (N',e')

The *Phi-Hiding Assumption* (ΦA) for c [10] states that (N,e) is computationally indistinguishable from (N',e') where (N,e) is generated via $\mathsf{RSA}_{inj}(1^k)$ and (N',e') is generated via $\mathsf{RSA}_{loss}(1^k)$. More precisely, to a distinguisher D we associate its *ΦA-advantage* defined as

$$\mathbf{Adv}^{\Phi\mathrm{A}}_{D,c}(k) \ =\ \Pr[\,D(N,e) \text{ outputs } 1\,] - \Pr[\,D(N',e') \text{ outputs } 1\,]$$

with inputs generated as above.

As shown by [25], RSA constitutes a *lossy trapdoor permutation* in the sense of [32] under *ΦA* by using the above two parameter generation algorithms. (We avoid giving a formal definition of lossy TDPs in the paper, since our results are specifically tied to ΦA.) We recall that we need $e \leq p^{1/2-\varepsilon}$ to avoid Coppersmith's attack [12,29] on ΦA. More specifically, N can be factored efficiently if $e \geq p^{1/2}$ and in time $O(N^\varepsilon)$ if $c = 1/4 - \varepsilon$ (*i.e.*, $\log e \geq \log N(1/4 - \varepsilon)$). For example, with modulus size $k = 2048$ we can set $\varepsilon = .04$ for 80-bit security (to enforce $k\varepsilon \geq 80$) and obtain $2048 \cdot (1/4 - 0.04) = 430$ bits of lossiness.

3 Approximate Regularity on Subdomains

We start by defining variants of regularity we consider.

NOTIONS OF REGULARITY ON SUBDOMAINS. Let $f\colon D \to R$ be a function from domain D to range R. We say that f is *regular* on D if $|f^{-1}(y)| = |D|/|R|$ for every $y \in R$. Suppose f is regular. Let $D' \subseteq D$ be a subdomain.

Definition 1 ($\mathcal{L}_1$-regularity). *We say that f is λ-$\mathcal{L}_1$-*regular on D' *if for all $y \in R$,*

$$\Delta(f(X), f(X')) \;\leq\; \lambda\,,$$

where $X \leftarrow\!\!\$\; D$ and $X' \leftarrow\!\!\$\; D'$.

Above "λ" is the approximation factor and "$\mathcal{L}_1$" indicates that we measure the regularity via the $\mathcal{L}_1$-norm.

We also consider the following "worst-case" regularity notion. For $y \in R$ we denote by $f^{-1}(y)[D']$ the preimage set of y restricted to D', that is

$$f^{-1}(y)[D'] := \{x \in D' \mid x \in f^{-1}(y)\}\,.$$

Definition 2 ($\mathcal{L}_\infty$-regularity). *We say that f is ν-$\mathcal{L}_\infty$-*regular on D' *if for all $y \in R$,*

$$\left| \frac{\left|f^{-1}(y)[D']\right|}{|D'|} - \frac{1}{|R|} \right| \;\leq\; \nu\,.$$

Equivalently, f is ν-$\mathcal{L}_\infty$-regular on D' if for all $y \in R$

$$|\Pr[f(X') = y \,:\, X' \leftarrow\!\!\$\; D'] - \Pr[f(X) = y \,:\, X \leftarrow\!\!\$\; D]| \;\leq\; \nu\,.$$

In other words, $\mathcal{L}_1$-regularity is a bound on the $\mathcal{L}_1$-distance of a random image point from the subdomain from uniform, and $\mathcal{L}_\infty$-regularity is a bound on the $\mathcal{L}_\infty$-distance from uniform. The following proposition (which immediately follows from the definitions) will be useful:

Proposition 3. *Suppose f is ν-$\mathcal{L}_\infty$-regular on D'. Then f is $\nu|R|$-$\mathcal{L}_1$-regular on D'.*

Thus, if f is ν-$\mathcal{L}_\infty$-regular on D' for $\nu \ll 1/|R|$, then f is $o(1)$-$\mathcal{L}_1$-regular on D'.

MAIN TECHNICAL QUESTION. We can now state the main (informal) technical question of this work. Consider the "lossy RSA" function $\mathsf{RSA}_{N',e'}$ where (N', e') is output by $\mathsf{RSA}_{loss}(1^k)$.

> What is the approximate regularity of $\mathsf{RSA}_{N',e'}$ on subdomains of $\mathbb{Z}_N$ of sufficient size?

In Section 4 we answer this question for a variety of parameter setting and regularity notions in the case that the subdomain of certain forms, in particular those described by *arithmetic progressions*. In Section 5 we give applications of these results to hardcore bits of RSA, RSA PKCS v1.5 encryption, and RSA-OAEP encryption.

4 Bounds on Approximate Regularity of Lossy RSA

We give bounds on the approximate regularity of lossy of RSA for a variety of parameter settings as notions of regularity.

4.1 $\mathcal{L}_1$-Regularity for Random Translations

We consider *expected* $\mathcal{L}_1$-regularity of lossy RSA over random translation of a fixed subset. The following lemma says for any subset of sufficient size, we have good expected $\mathcal{L}_1$-regularity over a random translation of the subset. It can also be viewed as saying that exponentiation with a random offset modulo N is a strong seeded extractor. (However, this interpretation is just for understanding; we do not use the lemma this way.)

Lemma 4. *Let $N = pq$ and e be such that $e \mid p-1$ and $\gcd(e, q-1) = 1$. Let $\mathcal{K} \subseteq \mathbb{Z}_N$ such that $|\mathcal{K}| \geq 4N/(e\alpha^2)$ for some $\alpha \geq \frac{4(p+q-1)}{N}$. Then*

$$(C, (C+X)^e \bmod N) \approx_\alpha (U', U^e \bmod N)$$

where $C, U', U \leftarrow\!\!\$\, \mathbb{Z}_N$ and $X \leftarrow\!\!\$\, \mathcal{K}$.

The proof relies on a careful collision probability argument. One key piece of that argument is the observation that the random offset and exponentiation *together* behave like a universal hash function: for any two values $a, b \in \mathbb{Z}_N$, the ratio $\left(\frac{a+C}{b+C}\right)^e$ is nearly uniformly distributed over e-th residues, conditioned on the denominator being invertible.

Proof. For ease of notation let $\mathcal{P}$ denote the distribution of $(C, (C+X)^e)$ and $\mathcal{U}$ denote the distribution of (C, U^e) (we omit the "mod N" here and below when it is clear from context). Let $K = |\mathcal{K}|$. Write

$$\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_0 \quad \text{and} \quad \mathcal{U} = \mathcal{U}_1 + \mathcal{U}_0$$

where $\mathcal{P}_1$ denotes the distribution of $\mathcal{P} \ \wedge \ (C+X)^e \in \mathbb{Z}_N^*$, $\mathcal{P}_0$ denotes the distribution of $\mathcal{P} \wedge (C+X)^e \notin \mathbb{Z}_N^*$, $\mathcal{U}_1$ denotes the distribution of $\mathcal{U} \wedge U^e \in \mathbb{Z}_N^*$, and $\mathcal{U}_0$ denotes the distribution of $\mathcal{U} \ \wedge \ U^e \notin \mathbb{Z}_N^*$. Note that

$$\Delta(\mathcal{P},\mathcal{U}) = ||\mathcal{P} - \mathcal{U}||_1 = ||\mathcal{P}_1 - \mathcal{U}_1||_1 + ||\mathcal{P}_0 - \mathcal{U}_0||_1 \ .$$

We bound each term with the following claims.

Claim.

$$||\mathcal{P}_1 - \mathcal{U}_1||_1 \leq \frac{\alpha}{2} \ .$$

Proof. We have

$$\begin{aligned} ||\mathcal{P}_1 - \mathcal{U}_1||_1 &\leq \sqrt{\mathrm{supp}(\mathcal{P}_1 - \mathcal{U}_1)} \cdot ||\mathcal{P}_1 - \mathcal{U}_1||_2 \\ &\leq \sqrt{\mathrm{supp}(\mathcal{P}_1 - \mathcal{U}_1) \cdot (||\mathcal{P}_1||_2)^2 - 1} \ . \end{aligned} \tag{1}$$

The first line above is by the Cauchy-Schwarz inequality, and the second is due to the fact that $(\mathcal{P}_1 - \mathcal{U}_1) \perp \mathbf{1}$, which follows from the observation that

$$\langle \mathbf{1}, \mathcal{P}_1 \rangle = \frac{p+q-1}{N} = \langle \mathbf{1}, \mathcal{U}_1 \rangle$$

where the first equality is because the marginal distribution of $(C+X)^e$ is that of U^e. To bound Equation (1), note that $\mathrm{supp}(\mathcal{P}_1 - \mathcal{U}_1) = N\phi(N)/e$. Also,

$$(||\mathcal{P}_1||_2)^2 = \Pr\left[\, (C,(C+X)^e) = (C',(C'+Y)^e \wedge z \in \mathbb{Z}_N^* \,\right] = \frac{1}{N} \cdot \Pr\left[\, (C+X)^e = (C+Y)^e \ \wedge z \in \mathbb{Z}_N^* \,\right]$$

where $X, Y \leftarrow\!\!\$\, \mathcal{K}$ and z denotes the common value of $(C+X)^e$ and $(C+Y)^e$. We have

$$\begin{aligned}
&\Pr\left[\, (C+X)^e = (C+Y)^e \ \wedge z \in \mathbb{Z}_N^* \,\right] = \sum_{\omega} \Pr\left[\, (C+X)/(C+Y) = \omega \ \wedge \ z \in \mathbb{Z}_N^* \,\right] \\
&\leq \Pr\left[\, X = Y \,\right] + \sum_{\omega \neq 1} \Pr\left[\, (C+X)/(C+Y) = \omega \ \wedge \ z \in \mathbb{Z}_N^* \mid X \neq Y \,\right] \cdot \Pr\left[\, X \neq Y \,\right] \\
&\leq \Pr\left[\, X = Y \,\right] + \sum_{\omega \neq 1} \Pr\left[\, C = (\omega Y - X)/(\omega - 1) \ \wedge \ z \in \mathbb{Z}_N^* \mid X \neq Y \,\right] \cdot \Pr\left[\, X \neq Y \,\right] \\
&\leq \frac{1}{K} + \frac{e-1}{N}\left(1 - \frac{1}{K}\right) .
\end{aligned}$$

where ω is an e-th root of unity modulo N (for which there are e possibilities); for the second-to-last line above we use the fact that $X = Y$ iff $\omega = 1$. Plugging the above into Equation (1) yields

$$\begin{aligned}
||\mathcal{P}_1 - \mathcal{U}_1||_1 &\leq \sqrt{\frac{\phi(N)}{e}\left(\frac{1}{K} + \frac{e-1}{N}\left(1 - \frac{1}{K}\right)\right) - 1} \\
&= \sqrt{\frac{\phi(N)/e}{K} + \left(\frac{e-1}{e} \cdot \frac{\phi(N)}{N} \cdot \frac{K-1}{K} - 1\right)} \ \leq \ \frac{\alpha}{2}
\end{aligned}$$

as desired, where for the last inequality we use the assumption $K \geq 4N/(e\alpha^2)$. ∎

Claim.

$$||\mathcal{P}_0 - \mathcal{U}_0||_1 \leq \frac{\alpha}{2} .$$

Proof. We have

$$||\mathcal{P}_0 - \mathcal{U}_0||_1 \ \leq \ \langle \mathbf{1}, \mathcal{P}_0 \rangle + \langle \mathbf{1}, \mathcal{U}_0 \rangle \ = \ \frac{2(p+q-1)}{N} \ \leq \ \frac{\alpha}{2}$$

where the last inequality uses the assumption that $\alpha \geq \frac{4(p+q-1)}{N}$. ∎∎

4.2 $\mathcal{L}_\infty$-Regularity for Arithmetic Progressions

We next consider the $\mathcal{L}_\infty$-regularity of lossy RSA on subdomains described by *arithmetic progressions*. We start with some definitions.

Arithmetic progressions. Recall that a subset $P \subseteq [1, N]$ is an *arithmetic progression* if it can be expressed as $P = \{\sigma\ell + \tau : 1 \leq \ell \leq K\}$ for some $\tau, \sigma \neq 0$. Here σ is called the *period* of the arithmetic progression.

Gauss sums. We define a *Gauss sum* as

$$\mathcal{G}_p(a, d) := \sum_{x=1}^{p} e_p(ax^d) \tag{2}$$

where $a, d \in \mathbb{N}$, $e(x) = e^{2\pi ix}$ and $e_p(x) := e(x/p)$. Trivially one has $|\mathcal{G}_p(a, d)| \leq p$. There are a variety of tighter bounds available for various choices of parameters which will be discussed later.

Connecting Gauss sums to Lossy RSA. First we show how estimates on Gauss sums imply results about approximate regularity of lossy RSA.

Lemma 5. *Let $N = pq$ and e be such that $e \mid p - 1$ and $\gcd(e, q - 1) = 1$. Assume that*

$$\max_{a \neq 0} \left| \mathcal{G}_p\left(a, \frac{p-1}{e}\right) \right| \leq Cp^\theta$$

for some $0 < \theta < 1$. Let $P = \{\sigma\ell + \tau : 1 \leq \ell \leq K\}$ ($\sigma, \tau \in \mathbb{N}$), $P \subseteq [1, N]$ denote an arithmetic progression with $(\sigma, N) = 1$, and let $\mathcal{K} = \{(x, N) = 1 : x \in P\}$. We then have that (assuming $p, q > 2$ and $|\mathcal{K}| \geq \max(p, q)$)

$$\left| \Pr_{x \leftarrow \mathcal{K}} [x^e = a] - \Pr_{x \leftarrow \mathbb{Z}_N^*} [x^e = a] \right| \leq 7|\mathcal{K}|^{-1} + 10Ce|\mathcal{K}|^{-1} p^{\theta - 1} \log(|\mathcal{K}|). \tag{3}$$

Proof. If a is not in the range of $x \mapsto x^e$ then $\Pr_{x \leftarrow \mathcal{K}} [x^e = a] = \Pr_{x \leftarrow \mathbb{Z}_N^*} [x^e = a] = 0$, so we assume that a is in the range. Next we recall some elementary number theory. We will identify an element $x \in \mathbb{Z}_m^*$ ($m = p, q$ or N) with its smallest positive integer representative which we will denote $\overline{x}$. The Chinese remainder theorem gives the isomorphism $\mathbb{Z}_N^* \cong \mathbb{Z}_p^* \oplus \mathbb{Z}_q^*$. This isomorphism is explicitly given from $\mathbb{Z}_N^*$ to $\mathbb{Z}_p^* \oplus \mathbb{Z}_q^*$ by the map $a \mapsto (\overline{a} \mod p, \overline{a} \mod q)$. Let $\mathcal{S} := \{x \in \mathbb{Z}_N^* : x^e = a\}$, $a_p = \overline{a} \mod p$, and $a_q = \overline{a} \mod q$. Denote by $\mathcal{S}_p := \{\overline{x} \mod p : x \in \mathcal{S}\} = \{u^e = a_p : u \in \mathbb{Z}_p^*\}$ and $\mathcal{S}_q := \{\overline{x} \mod q : x \in \mathcal{S}\} = \{v^e = a_q : v \in \mathbb{Z}_q^*\}$. Since $(e, q - 1) = 1$ we have that that map $v \mapsto v^e$ is a bijection on $\mathbb{Z}_q^*$ and hence $|\mathcal{S}_q| = 1$. We will denote this element as s_q. The map $u \mapsto u^e$ on $\mathbb{Z}_p^*$ is e-to-1, so $|\mathcal{S}_p| = e = |\mathcal{S}|$. Moreover, $\mathcal{S}_p$ is a coset of a subgroup and can be represented as $\mathcal{S}_p = \{x^{\frac{p-1}{e}} b : x \in \mathbb{Z}_p^*\}$, for any $b \in \mathcal{S}_p$. Our goal is to estimate $|\mathcal{K} \cap \mathcal{S}|$. Given a set $S \subseteq \mathbb{Z}_m^*$ (for $m = p, q$ or N), we will denote the associated indicator function as $1_S(x) \mapsto \{0, 1\}$. Thus,

$$|\mathcal{K} \cap \mathcal{S}| = \sum_{x \in \mathcal{K}} 1_{\mathcal{S}}(x) = \sum_{x \in \mathcal{K}} 1_{s_q}(x) 1_{\mathcal{S}_p}(x) = \sum_{\substack{x \in \mathcal{K} \\ x \equiv s_q \mod q}} 1_{\mathcal{S}_p}(x).$$

We can expand $1_{\mathcal{S}_p}(x) = \sum_{\xi \in \mathbb{Z}_p^*} \widehat{1_{\mathcal{S}_p}}(\xi) e_p(x\xi)$ where the Fourier coefficients $\widehat{1_{\mathcal{S}_p}}(\xi)$ are given by

$$\widehat{1_{\mathcal{S}_p}}(\xi) = p^{-1} \sum_{x \in \mathbb{Z}_p} 1_{\mathcal{S}_p}(x) e_p(-x\xi).$$

We have that

$$\widehat{\mathcal{S}_p}(\xi) = p^{-1} \frac{e}{p-1} \mathcal{G}(b\xi, \frac{p-1}{e}).$$

Thus $\widehat{1_{\mathcal{S}_p}}(0) = \frac{e}{p-1}$ and $|\widehat{\mathcal{S}_p}(\xi)| \leq C \frac{e}{p-1} p^{\theta-1}$ for $\xi \neq 0$.

Let $\mathcal{K}' := \{\overline{x} \equiv s_q \mod q : x \in \mathcal{K}\}$. We wish to estimate $|\mathcal{K}'|$ in terms of $|\mathcal{K}|$. In what follows we make use of the assumption that $(\sigma, N) = 1$ which prevents P from degenerating to a point when reduced $\mod p$ or $\mod q$. Noting that $\mathcal{K}$ is obtained from P by sieving out elements congruent to $0 \mod p$ and $0 \mod q$, we have that $|P|(1 - p^{-1} - q^{-1}) + \mathcal{O}(3) = |\mathcal{K}|$. Now if we define $I := \{\overline{x} \equiv s_q \mod q : x \in P\}$ and $E := \{\overline{x} \equiv 0 \mod p : x \in I\}$, then $\mathcal{K}' = I \setminus E$. Moreover, $|E| \leq 1$ since if $b \in E \subseteq \mathbb{Z}_N^*$ then $b \equiv s_q \mod q$ and $b \equiv 0 \mod p$ which uniquely specifies b by the Chinese remainder theorem. Thus $|\mathcal{K}'| = |P| q^{-1} + \mathcal{O}(2)$ where $|P| = (|\mathcal{K}| + \mathcal{O}(3)) \frac{qp}{qp-p-q}$, which gives

$$|\mathcal{K}'| = (|\mathcal{K}| + \mathcal{O}(3)) \frac{p}{qp-p-q} + \mathcal{O}(2) = \frac{1}{q-1} |\mathcal{K}| + \mathcal{O}(7)$$

where we have used that $\frac{p}{qp-p-q} \leq 1$ (using that $p, q > 2$) and $\frac{p}{qp-p-q} = \frac{1}{q-1-qp^{-1}} = \frac{1}{q-1} + \frac{q-1+qp^{-1}}{(q-1)(q-1-qp^{-1})} = \frac{1}{q-1} + \mathcal{O}(2)$.

We now may express

$$1_{\mathcal{S}_p}(x) = \frac{e}{p-1} + \sum_{\xi \in \mathbb{Z}_p^*} \widehat{1_{\mathcal{S}_p}}(\xi) e_p(x\xi).$$

Thus

$$|\mathcal{K} \cap \mathcal{S}| = \sum_{x \in I \setminus E} 1_{\mathcal{S}_p}(\overline{x}) = \frac{e|\mathcal{K}'|}{(p-1)} + \sum_{\xi \in \mathbb{Z}_p^*} \widehat{1_{\mathcal{S}_p}}(\xi) \sum_{x \in I \setminus E} e_p(x\xi)$$

Using that $|\widehat{\mathcal{S}_p}(\xi)| \leq Cep^{\theta-2}$ we have

$$\left| |\mathcal{K} \cap \mathcal{S}| - \frac{e|\mathcal{K}|}{\phi(N)} \right| \leq 7 + C \frac{e}{p-1} p^{\theta-1} \sum_{\xi \in \mathbb{Z}_p^*} \left| \sum_{x \in I \setminus E} e_p(x\xi) \right|$$

Now I can be expressed as an arithmetic progression, $I = \{xq + b : x = 1, 2, \ldots |I|\}$ so

$$|\sum_{x \in I} e_p(x\xi)| = |\sum_{x=1}^{|I|} e_p(-b\xi) e_p(qx\xi)| = |\sum_{x=1}^{|I|} e_p(qx\xi)| = \left| \frac{\sin(\pi\xi q |I|/p)}{\sin(\pi\xi q/p)} \right| \leq \left| \frac{1}{\sin(\pi q \xi/p)} \right|.$$

Using the inequality $\sin(\pi x) \geq 2x$ for $-1/2 \leq x \leq 1/2$ and denoting the distance of a real number to the nearest integer by $||\cdot||$ the quantity above is $\leq 2^{-1}||\xi q p^{-1}||^{-1}$. Thus,

$$\sum_{\xi \in \mathbb{Z}_p^*} \left| \sum_{x \in I \setminus E} e_p(x\xi) \right| \leq 1 + 2^{-1} \sum_{\xi \in \mathbb{Z}_p^*} ||\xi q p^{-1}||^{-1} \leq 1 + 2^{-1} p \sum_{n=1}^{|I|} n^{-1} \leq 1 + 2^{-1} p (1 + \log(|I|))$$

where we have used the inequality $\sum_{n=1}^{|I|} n^{-1} \leq 1 + \log(|I|)$. Since $\Pr_{x \leftarrow \mathbb{Z}_N^*}[x^e = a] = \frac{e}{\phi(N)}$ and $\Pr_{x \leftarrow \mathcal{K}}[x^e = a] = \frac{|\mathcal{K} \cap \mathcal{S}|}{|\mathcal{K}|}$, putting this all together we have that

$$\left| \Pr_{x \leftarrow \mathcal{K}}[x^e = a] - \Pr_{x \leftarrow \mathbb{Z}_N^*}[x^e = a] \right| \leq 7|\mathcal{K}|^{-1} + C \frac{e}{p-1} |\mathcal{K}|^{-1} p^{\theta - 1} \left(1 + 2^{-1} p \left(\log \left(\frac{|\mathcal{K}|}{q} + 1 \right) + 1 \right) \right).$$

From our assumptions we have $\log\left(\frac{|\mathcal{K}|}{q} + 1\right) \leq \log(|\mathcal{K}|)$, and $\log\left(\frac{|\mathcal{K}|}{q} + 1\right) + 1 \leq 2\log(|\mathcal{K}|)$. Now $1 + 2^{-1} p \times 2\log(|\mathcal{K}|) \leq 5\log(|\mathcal{K}|)$. Using $\frac{1}{p-1} \leq \frac{2}{p}$, we have

$$\left| \Pr_{x \leftarrow \mathcal{K}}[x^e = a] - \Pr_{x \leftarrow \mathbb{Z}_N^*}[x^e = a] \right| \leq 7|\mathcal{K}|^{-1} + 10Ce|\mathcal{K}|^{-1} p^{\theta - 1} \log(|\mathcal{K}|).$$

This completes the proof. ∎

Known estimates on Gauss sums. We now summarize some known estimates on $\mathcal{G}_p(a, e)$. Throughout, we assume $1 \leq a < p$.

$$|\mathcal{G}_p(a,d)| \leq \begin{cases} (d-1)p^{1/2} & , 1 \leq d \leq p^{1/3} \\ 2 \cdot 3^{-1/4} d^{5/8} p^{5/8} & , p^{1/3} < d \leq p^{1/2} \\ 2 \cdot 3^{-1/4} d^{3/8} p^{3/4} & , p^{1/2} < d \leq p^{2/3} \\ C_\delta p^{1 - \varepsilon(\delta)} & , p^{2/3} < d < p^\delta. \end{cases}$$

The first estimate is classical, the second and third are due to Heath-Brown and Konyagin [17] (with the explicit constants given by Cochrane and Pinner [11]), and the fourth is due to Bourgain, Glibichuk, and Konyagin [8]. To the best of our knowledge explicit values of C_δ have not been worked out. Also, see [27] and [7] for some additional refinements. Much more is believed to be true, in particular Montgomery, Vaughan, and Wooley [30] have made the following conjecture

$$|\mathcal{G}_p(a,d)| \leq \min\{(d-1)p^{1/2}, (1+\eta)(2dp\log(dp))^{1/2}\} \tag{4}$$

where $\eta \to 0$ as d and p/d tend to infinity.

Bounds on regularity of Lossy RSA. Combining the known estimates with Lemma 5 gives the following corollary.

Corollary 6. *With the notation and assumptions of Lemma 5 we have*

$$\text{The map } x \mapsto x^e \text{ is } \nu\text{-}\mathcal{L}_\infty\text{-regular on } \mathcal{K} \text{ for } \begin{cases} \nu = C_\delta e p^{-\varepsilon(\delta)} \frac{\log(|\mathcal{K}|)}{|\mathcal{K}|} & , p^\delta \le e \le p^{1/3} \\ \nu = 23 e^{5/8} p^{1/8} \frac{\log(|\mathcal{K}|)}{|\mathcal{K}|} & , p^{1/3} < e \le p^{1/2} \\ \nu = 23 e^{3/8} p^{1/4} \frac{\log(|\mathcal{K}|)}{|\mathcal{K}|} & , p^{1/2} < e \le p^{2/3} \\ \nu = 17 p^{1/2} \frac{\log(|\mathcal{K}|)}{|\mathcal{K}|}. & , p^{2/3} < e < p. \end{cases}$$

Recall from Section 2 that we take $e \le p^{1/2-\varepsilon}$ in our applications, and thus only the first two cases above are applicable. We include bounds for other parameter ranges in case they are useful for future work. (Indeed, we consider finding a relaxation of lossiness that avoids Coppersmith's attack but still allows our proof techniques to go through to be an interesting open problem.) We are also able to obtain improvements for some values of e and K by appealing to estimates for incomplete character sums, see the full version [28] for details.

Consequences of the MVW conjecture and comparison to collision probability bound. One may check that the Montgomery-Vaughan-Wooley (MVW) conjecture mentioned above would give ν-$\mathcal{L}_\infty$-regularity for $\nu = O\left(e^{1/2} \log^{1/2}(p) \frac{\log(|\mathcal{K}|)}{|\mathcal{K}|}\right)$.[4] Thus, disregarding logarithmic factors the MWV conjecture implies λ-$\mathcal{L}_1$-regularity for $\lambda = O(\sqrt{N/eK} \cdot \sqrt{N/K})$, whereas the collision probability bound in Section 4.1 gives $\lambda = O(\sqrt{N/eK})$, which is which is always better since $K \le N$. However, when $K \sim N$ these bounds essentially agree and are both asymptotically around $1/\sqrt{e}$.

5 Applications

We describe several applications of our results.

5.1 Large Consecutive Runs of Hardcore Bits for RSA

We can use our result on expected $\mathcal{L}_1$-regularity of lossy RSA over random translations given in Section 4.1 to derive new results on substrings of the RSA input that are hardcore.

Hardcore substrings. We begin by defining what it means for a substring of the input to be hardcore. For $1 \le i < j \le k$, we say that the (i,j)-*th substring of RSA is simultaneously hardcore* (we omit "simultaneously" below, with it being understood) if the following two distributions are computationally indistinguishable:

$$\mathsf{DistReal} := \{(N, e, x^e \bmod N, x[i \ldots j])) \ : \ (N,e) \leftarrow\!\!\$\ \mathsf{RSA}_{inj}(1^k)\ ;\ x \leftarrow\!\!\$\ \mathbb{Z}_N^*\}$$

$$\mathsf{DistRand} := \{(N, e, x^e \bmod N, r) \ : \ (N,e) \leftarrow\!\!\$\ \mathsf{RSA}_{inj}(1^k)\ ;\ x \leftarrow\!\!\$\ \mathbb{Z}_N^*\ ;\ r \leftarrow\!\!\$\ \{0,1\}^{j-i}\}\ .$$

[4] Note that Parseval's identity gives us that $\sum_{a \in \{x^{(p-1)/d} : x \in Z_p\}} |\mathcal{G}_p(a,d)|^2 \gg d^2 p$ (see section 4 of [30]). Thus, for some a we must have $|\mathcal{G}(p,a)| \gg (dp)^{1/2}$. So (disregarding logarithmic factors) nothing beyond the MWV conjecture is possible.

To a distinguisher D we associate its *hardcore-bits advantage* defined as

$$\mathbf{Adv}^{\mathrm{hcb}}_{i,j,D}(k) \;=\; \Pr[\, D(\mathsf{DistReal}) \text{ outputs } 1\,] - \Pr[\, D(\mathsf{DistRand}) \text{ outputs } 1\,]\,.$$

Our result. We are now ready to state our result: Roughly, under ΦA, (1) the $\log e - 3\log(1/\varepsilon)$ most significant bits of RSA are hardcore, (2) the $\log e - 3\log(1/\varepsilon)$ least significant bits of RSA are hardcore, and (3) an arbitrary substring of $\log e - 4\log(1/\varepsilon)$ bits of RSA are hardcore. Typically, in practice the least significant bits are used, see e.g. [39].

Theorem 7. *Assume that ΦA holds for c and let $\varepsilon > 0$. Let $1 \le i < j \le n$ such that $|j - i| \le \log e - 4\log(1/\varepsilon) - 2$. Then for any hardcore-bits distinguisher D against RSA there is a ΦA distinguisher D' such that for all $k \in \mathbb{N}$*

$$\mathbf{Adv}^{\mathrm{hcb}}_{i,j,D}(k) \le \mathbf{Adv}^{\Phi\mathrm{A}}_{c,D'}(k) + 2\varepsilon\,.$$

The running-time of D is that of D'. In the special cases $i = 1$ or $j = n$ (i.e., for the least significant or most significant bits), we only require $|j - i| \le \log e - 3\log(1/\alpha) - 2$.

The proof is in the full version [28]. The main idea is to note that, in Lemma 4, if we choose X appropriately, the (i, j)-th substrings of X and $X + C$ in the lemma will be the same. To ensure this we need to choose X so that, with high probability, we avoid "overflow" modulo N, or a carry into the the i-th bit position in the addition. Ensuring this is why we pay an extra $2\log(1/\varepsilon)$ bits (versus Lemma 4) in the theorem in general.

Concrete parameters. Recall from Section 2 that with modulus size $k = 2048$ we can take $\log e$ to be roughly 430 bits for 80-bit security. Then, taking $\varepsilon = 2^{-80}$ in Theorem 7, we get 190 natural hardcore bits of RSA (either the 190 most significant bits or the 190 least significant bits). Similarly, with $k = 3072$ we get 688 bits of lossiness and 448 natural hardcore bits.

5.2 IND-CPA Security of PKCS #1 v1.5

We can use our results on $\mathcal{L}_\infty$-regularity of lossy RSA on arithmetic progressions given in Section 4.2 to prove security of PKCS #1 v1.5 encryption.

PKCS #1 v1.5 encryption. Namely, define the "simple embedding" RSA-based encryption scheme defined as follows. Let μ, ρ such that $\mu + \rho + 32 = k$ be integer parameters. Define the randomized encoding function PKCS that takes plaintext $x \in \{0,1\}^\mu$ and coins $r \in \{0,1\}^\rho$ and outputs

$$\mathsf{PKCS}(x; r) = x \| 00_{16} \| r\,.$$

Define the encryption scheme $\Pi_{\mathsf{PKCS}} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ by

Alg $\mathsf{Kg}(1^k)$	**Alg** $\mathsf{Enc}((N,e),x)$	**Alg** $\mathsf{Dec}((N,d),y)$
$(N,e,d) \leftarrow\!\!\$\ \mathsf{RSA}_{inj}(1^k)$	$x' \leftarrow\!\!\$\ \mathsf{PKCS}(x)$	$x' \leftarrow y^d \bmod N$
Return $((N,e),(N,d))$	$x'' \leftarrow 0002_{16} \| x'$	$0002_{16} \| x \| 00_{16} \| r \leftarrow x'$
	$y \leftarrow (x'')^e \bmod N$	Return x
	Return y	

Essentially such a scheme was adopted by PKCS #1 v1.5 and is still in widespread use. (In practice, as opposed to in the academic literature, r and x are switched; however, this doesn't affect our results.)

Our result. We show the first positive result about the security of this scheme; namely, for certain parameters it is IND-CPA secure. (The standard definition of IND-CPA security is recalled in Appendix A.)

Theorem 8. *Suppose ΦA holds for c and lossy RSA is ν-$\mathcal{L}_\infty$-regular on arithmetic progressions of length 2^ρ. Then for any IND-CPA adversary A against Π_{PKCS}, there is a distinguisher D against ΦA such that for every $k \in N$*

$$\mathbf{Adv}^{\text{ind-cpa}}_{\Pi_{\mathsf{PKCS}},A}(k) \;\le\; \mathbf{Adv}^{\Phi\text{A}}_{c,D}(k) + \frac{\nu \cdot \phi(N)}{e} \;.$$

The running-time of D is that of A.

Proof. (Sketch.) The first step of the proof is to replace (N,e) generated via $\mathsf{RSA}_{inj}(1^k)$ in the IND-CPA game with (N',e') generated via $\mathsf{RSA}_{loss}(1^k)$. Now consider the distribution of a ciphertext $C = (0002_{16} \| x \| 00_{16} \| R)^{e'} \bmod N'$ for any fixed but arbitrary plaintext $x \in \{0,1\}^\mu$, where $R \in \{0,1\}^\rho$ is uniformly random. (Note that x may depend on N', e' here, which are fixed in the argument below.) Notice subdomain $\{0002_{16} \| x \| 00_{16} \| r \mid r \in \{0,1\}^\rho\}$ is described by the arithmetic progression of length 2^ρ, namely $\{2^{\rho+16+\mu} \cdot 2 + 2^{\rho+16} \cdot x + i \mid 1 \le i \le 2^\rho\}$. (Variants of the scheme, e.g. where R and x are switched, are described by an arithmetic progression with a different period.) Proposition 3 tells us that lossy RSA is $\nu(\phi(N)/e)$-$\mathcal{L}_1$-regular on arithmetic progressions of length 2^ρ, and thus $\Delta(C, U^e) \le \nu(\phi(N)/e)$. Noting that U^e is independent of x concludes the proof. ∎

Concrete parameters. We can calculate concrete security bounds for the scheme according to our results on the $\mathcal{L}_\infty$-regularity of lossy RSA on arithmetic progressions given in Section 4.2. As per Section 2 assume $e = p^{1/2-\varepsilon}$. Then according to part 2 of Corollary 6 we have that lossy RSA is ν-$\mathcal{L}_\infty$-regular on arithmetic progressions of length K for

$$\nu = 23 e^{5/8} p^{1/8} \frac{\log(K)}{K} \;.$$

Now $p = N^{1/2}$ so $e = N^{1/4-\varepsilon/2}$ and thus

$$\nu \;=\; \frac{23 \cdot N^{7/32-(5/16)\varepsilon} \log K}{K} \;=\; \frac{e}{\phi(N)} \left(\frac{N}{e} \cdot \frac{23 \cdot N^{7/32-(5/16)\varepsilon} \log K}{K} \right)$$

$$= \frac{e}{\phi(N)} \left(\frac{23 \cdot N^{31/32+(3/16)\varepsilon} \log K}{K} \right)$$

Imposing

$$\nu \leq \frac{e}{\phi(N)} N^{-\varepsilon} \implies K \geq 23 \cdot N^{31/32+11/8\varepsilon} \cdot \log N \ .$$

Unfortunately, we must use a very long modulus length to get any meaningful concrete bound. For example, consider modulus length $k = 8192$. By taking $\varepsilon = .01$ we get 80-bit security and approximately 128-bit messages. Hence, we view our result as a qualitative one (in the same vein, we note, for example, that the current concrete security reduction for RSA-OAEP requires modulus length 4096 [33] for a meaningful bound) and hope further work will improve the parameters.

5.3 Improved Security Reduction for RSA-OAEP

Finally, we can use our bounds on $\mathcal{L}_\infty$-regularity of lossy RSA on arithmetic progressions given in Section 4.2 to improve the bounds given in [25] on CPA security of RSA-OAEP encryption scheme [4], as adopted by PKCS #1 v2.1 as a replacement for the "simple embedding" scheme described above.

RSA-OAEP ENCRYPTION. Let μ, ρ such that $\mu + \rho + 16 = k$ be integer parameters. Define the randomized encoding function OAEP that takes plaintext $x \in \{0,1\}^\mu$ and coins $r \in \{0,1\}^\rho$ and outputs

$$\mathsf{OAEP}(x; r) = G(r) \oplus x \| H(s) \oplus r$$

where $s = G(r) \oplus x$ and G, H are hash functions. Define scheme $\Pi_{\mathsf{OAEP}} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ by

Alg $\mathsf{Kg}(1^k)$	**Alg** $\mathsf{Enc}((N,e),x)$	**Alg** $\mathsf{Dec}((N,d),y)$
$(N,e,d) \leftarrow\!\!\$\ \mathsf{RSA}_{inj}(1^k)$	$x' \leftarrow\!\!\$\ 0002_{16} \| \mathsf{OAEP}(x)$	$x' \leftarrow y^d \bmod N$
Return $((N,e),(N,d))$	$y \leftarrow (x')^e \bmod N$	$0002_{16} \| s \| t$
	Return y	$r \leftarrow t \oplus H(s)$; $x \leftarrow s \oplus G(r)$
		Return x

OUR RESULT. Recall that Theorem 4.2 of [25] provides several bounds on the CPA security of RSA-OAEP. The best bound, namely Part 2 of Theorem 4.2, requires that RSA is (close to) regular on the subdomain where the most significant two bytes of the input are zero. However, they left this as an open problem and resorted to a worse bound for general functions, namely Part 3 of Theorem 4.2. Here we adapt Part 2 to prove our result:

Theorem 9. *Suppose ΦA holds and $\mathsf{RSA}_{N',e'}$ as defined above is $\lambda(e/\phi(N))$-$\mathcal{L}_\infty$-regular on arithmetic progressions of length 2^{k-16}, and G is t-wise independent. Then for every IND-CPA adversary A against Π_{OAEP} there is a distinguisher D against ΦA such that*

$$\mathbf{Adv}^{\text{ind-cpa}}_{\Pi_{\mathsf{OAEP}},A}(k) \ \leq \ \mathbf{Adv}^{\Phi\text{A}}_{c,D}(k) + 2^{-u}$$

where

$$u = \frac{t}{2t+2}(\log \lambda + \rho - s - \log t + 2) - \frac{\mu + s + 2}{t+1} - 1$$

The running-time of D' is that of D.

Specifically, "$\log \lambda$" term in the above theorem was absent in Part 2 of Theorem 4.2 [25], but its presence here follows from their analysis (see "Proof of Part 1", pg. 14, of [25]).

Concrete parameters. Notice that, in this application, as opposed to Section 5.2, we only need regularity on very large arithmetic progressions of length 2^{k-16} (independent of μ, ρ) versus length 2^{ρ} in Section 5.2. More precisely, we need ν-$\mathcal{L}_\infty$-regularity for say $\nu \leq e/\phi(N) \cdot 1/2$ (rather than $\nu \leq e/\phi(N) \cdot \mathsf{negl}(k)$) to essentially lose nothing in the bound as compared to the bound of Part 2 of Theorem 4.1 in [25] (we lose just $\log 2 = 1$ additional bit). Following our calculations in Section 5.2 we impose

$$\nu \leq \frac{e}{\phi(N)} N^{-\delta} \implies K \geq 23 \cdot N^{31/32+3/8\varepsilon+\delta} \cdot \log N \ .$$

For modulus length $k = 2048$ we take take $\varepsilon = .04$ and $\delta = .001$ for 80-bit security, and obtain $\log K \geq 2032$, meaning we can indeed fix 16 bits of the domain to zeros for ν-regularity. The concrete value of these of this savings is significant. For example, with $k = 2048$ we can support 274-bit messages for 80-bit security rather than 160-bits as obtained by the general bound of [25, Part 1, Theorem 4.2].

Acknowledgements. We thank the Eurocrypt 2013 anonymous reviewers for their helpful comments. A.O. is supported in part by NSF grants CNS-1012910 and CNS-0546614; additionally, this work was done in part while at the University of Texas at Austin, supported by NSF grants CNS-0915361 and CNS-0952692. M.L. is supported by a NSF Postdoctoral Fellowship, DMS-1204206.

References

1. Bellare, M., Brakerski, Z., Naor, M., Ristenpart, T., Segev, G., Shacham, H., Yilek, S.: Hedged public-key encryption: How to protect against bad randomness. In: Matsui, M. (ed.) ASIACRYPT 2009. LNCS, vol. 5912, pp. 232–249. Springer, Heidelberg (2009)
2. Bellare, M., Hofheinz, D., Yilek, S.: Possibility and impossibility results for encryption and commitment secure under selective opening. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 1–35. Springer, Heidelberg (2009)
3. Bellare, M., Rogaway, P.: Random oracles are practical: A paradigm for designing efficient protocols. In: Ashby, V. (ed.) ACM CCS 1993, pp. 62–73. ACM Press (November 1993)
4. Bellare, M., Rogaway, P.: Optimal asymmetric encryption. In: De Santis, A. (ed.) EUROCRYPT 1994. LNCS, vol. 950, pp. 92–111. Springer, Heidelberg (1995)
5. Bleichenbacher, D.: Chosen ciphertext attacks against protocols based on the RSA encryption standard PKCS #1. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 1–12. Springer, Heidelberg (1998)
6. Boldyreva, A., Fehr, S., O'Neill, A.: On notions of security for deterministic encryption, and efficient constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 335–359. Springer, Heidelberg (2008)

7. Bourgain, J.: Multilinear exponential sums in prime fields under optimal entropy condition on the sources. Geom. Funct. Anal. 18(5), 1477–1502 (2009)
8. Bourgain, J., Glibichuk, A.A., Konyagin, S.V.: Estimates for the number of sums and products and for exponential sums in fields of prime order. J. London Math. Soc. (2) 73(2), 380–398 (2006)
9. Cachin, C.: Efficient private bidding and auctions with an oblivious third party. In: ACM CCS 1999, pp. 120–127. ACM Press (November 1999)
10. Cachin, C., Micali, S., Stadler, M.A.: Computationally private information retrieval with polylogarithmic communication. In: Stern, J. (ed.) EUROCRYPT 1999. LNCS, vol. 1592, pp. 402–414. Springer, Heidelberg (1999)
11. Cochrane, T., Pinner, C.: Explicit bounds on monomial and binomial exponential sums. Q. J. Math. 62(2), 323–349 (2011)
12. Coppersmith, D.: Small solutions to polynomial equations, and low exponent RSA vulnerabilities. Journal of Cryptology 10(4), 233–260 (1997)
13. Coron, J.-S., Joye, M., Naccache, D., Paillier, P.: New attacks on PKCS#1 v1.5 encryption. In: Preneel, B. (ed.) EUROCRYPT 2000. LNCS, vol. 1807, pp. 369–381. Springer, Heidelberg (2000)
14. Gentry, C., Mackenzie, P.D., Ramzan, Z.: Password authenticated key exchange using hidden smooth subgroups. In: Atluri, V., Meadows, C., Juels, A. (eds.) ACM CCS 2005, pp. 299–309. ACM Press (November 2005)
15. Goldreich, O.: Foundations of Cryptography: Basic Applications, vol. 2. Cambridge University Press, Cambridge (2004)
16. Goldwasser, S., Micali, S.: Probabilistic encryption. Journal of Computer and System Sciences 28(2), 270–299 (1984)
17. Heath-Brown, D.R., Konyagin, S.: New bounds for Gauss sums derived from kth powers, and for Heilbronn's exponential sum. Q. J. Math. 51(2), 221–235 (2000)
18. Hemenway, B., Ostrovsky, R.: Public-key locally-decodable codes. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 126–143. Springer, Heidelberg (2008)
19. Hofheinz, D., Kiltz, E.: Practical chosen ciphertext secure encryption from factoring. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 313–332. Springer, Heidelberg (2009)
20. Hohenberger, S., Waters, B.: Short and stateless signatures from the RSA assumption. In: Halevi, S. (ed.) CRYPTO 2009. LNCS, vol. 5677, pp. 654–670. Springer, Heidelberg (2009)
21. Jager, T., Kohlar, F., Schäge, S., Schwenk, J.: On the security of TLS-DHE in the standard model. In: Safavi-Naini, R. (ed.) CRYPTO 2012. LNCS, vol. 7417, pp. 273–293. Springer, Heidelberg (2012)
22. Jager, T., Schinzel, S., Somorovsky, J.: Bleichenbacher's attack strikes again: Breaking PKCS#1 v1.5 in XML encryption. In: Foresti, S., Yung, M., Martinelli, F. (eds.) ESORICS 2012. LNCS, vol. 7459, pp. 752–769. Springer, Heidelberg (2012)
23. Kakvi, S.A., Kiltz, E.: Optimal security proofs for full domain hash, revisited. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 537–553. Springer, Heidelberg (2012)
24. Katz, J., Lindell, Y.: Introduction to Modern Cryptography. Chapman and Hall/CRC Press (2007)
25. Kiltz, E., O'Neill, A., Smith, A.: Instantiability of RSA-OAEP under chosen-plaintext attack. In: Rabin, T. (ed.) CRYPTO 2010. LNCS, vol. 6223, pp. 295–313. Springer, Heidelberg (2010)
26. Kohel, D.R., Shparlinski, I.: On exponential sums and group generators for elliptic curves over finite fields. In: Bosma, W. (ed.) ANTS 2000. LNCS, vol. 1838, pp. 395–404. Springer, Heidelberg (2000)

27. Konyagin, S.V.: Estimates for trigonometric sums over subgroups and for Gauss sums. In: IV International Conference "Modern Problems of Number Theory and its Applications": Current Problems, Part III (Russian) (Tula, 2001), pp. 86–114. Mosk. Gos. Univ. im. Lomonosova, Mekh.-Mat. Fak., Moscow (2002)
28. Lewko, M., O'Neill, A., Smith, A.: Regularity of lossy RSA on subdomains and its applications. Cryptology ePrint Archive (2013), http://eprint.iacr.org/
29. May, A.: Using lll-reduction for solving rsa and factorization problems: A survey. In: LLL+25 Conference in Honour of the 25th Birthday of the LLL Algorithm (2007), http://www.cits.rub.de/imperia/md/content/may/paper/lll.ps
30. Montgomery, H.L., Vaughan, R.C., Wooley, T.D.: Some remarks on Gauss sums associated with kth powers. Math. Proc. Cambridge Philos. Soc. 118(1), 21–33 (1995)
31. Nishimaki, R., Fujisaki, E., Tanaka, K.: Efficient non-interactive universally composable string-commitment schemes. In: Pieprzyk, J., Zhang, F. (eds.) ProvSec 2009. LNCS, vol. 5848, pp. 3–18. Springer, Heidelberg (2009)
32. Peikert, C., Waters, B.: Lossy trapdoor functions and their applications. In: Ladner, R.E., Dwork, C. (eds.) 40th ACM STOC, pp. 187–196. ACM Press (May 2008)
33. Pointcheval, D.: How to encrypt properly with rsa. RSA Laboratories Crypto Bytes 5(1), 9–19 (2002)
34. Rivest, R.L., Shamir, A., Adleman, L.M.: A method for obtaining digital signature and public-key cryptosystems. Communications of the Association for Computing Machinery 21(2), 120–126 (1978)
35. Schridde, C., Freisleben, B.: On the validity of the Φ-hiding assumption in cryptographic protocols. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 344–354. Springer, Heidelberg (2008)
36. Shoup, V.: A proposal for an ISO standard for public-key encryption. ISO/IEC JTC (2001), http://www.shoup.net/papers/iso-2_1.pdf
37. Shparlinski, I.: On gaussian sums for finite fields and elliptic curves. In: Lobstein, A., Litsyn, S.N., Zémor, G., Cohen, G. (eds.) Algebraic Coding 1991. LNCS, vol. 573, pp. 5–15. Springer, Heidelberg (1992)
38. Shparlinski, I.: Bilinear character sums over elliptic curves. Finite Fields and Their Applications 14(1), 132–141 (2008)
39. Steinfeld, R., Pieprzyk, J., Wang, H.: On the provable security of an efficient RSA-based pseudorandom generator. In: Lai, X., Chen, K. (eds.) ASIACRYPT 2006. LNCS, vol. 4284, pp. 194–209. Springer, Heidelberg (2006)

A Public-key Encryption and Its Security

A *public-key encryption scheme* with message-space $MsgSp$ is a triple of algorithms $\Pi = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$. The key-generation algorithm Kg returns a public key pk and matching secret key sk. The encryption algorithm Enc takes pk and a plaintext m to return a ciphertext. The deterministic decryption algorithm Dec takes sk and a ciphertext c to return a plaintext. We require that for all messages $m \in MsgSp$

$$\Pr[\mathsf{Dec}(sk, \mathsf{Enc}(pk, m)) \neq m \ : \ (pk, sk) \leftarrow\!\!\$\ \mathsf{Kg}]$$

is negligible.

To an encryption scheme $\Pi = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ and an adversary $A = (A_1, A_2)$ we associate a chosen-plaintext attack experiment,

$$\begin{array}{l} \textbf{Experiment } \mathbf{Exp}^{\text{ind-cpa}}_{\Pi,A}(k) \\ \quad b \leftarrow\!\!\$\ \{0,1\}\ ;\ (pk, sk) \leftarrow\!\!\$\ \mathsf{Kg}(1^k) \\ \quad (m_0, m_1, St) \leftarrow\!\!\$\ A_1(pk) \\ \quad c \leftarrow\!\!\$\ \mathsf{Enc}(pk, m_b) \\ \quad d \leftarrow\!\!\$\ A_2(pk, c, St) \\ \quad \text{If } d = b \text{ then return 1 else return 0} \end{array}$$

where we require A's output to satisfy $|m_0| = |m_1|$. Define the *ind-cpa advantage* of A against Π as

$$\mathbf{Adv}^{\text{ind-cpa}}_{\Pi,A}(k) = 2 \cdot \Pr\left[\, \mathbf{Exp}^{\text{ind-cpa}}_{\Pi,A}(k) \text{ outputs } 1 \,\right] - 1 \ .$$