On the Regularity of Lossy RSA: Improved Bounds and Applications to Padding-Based Encryption

Adam Smith* Ye Zhang*

October 6, 2014

Abstract

We provide new bounds on how close to regular the map $x \mapsto x^e$ is on arithmetic progressions in $\mathbb{Z}_N$, assuming $e|\Phi(N)$ and N is composite. We use these bounds to analyze the security of natural cryptographic problems related to RSA, based on the well-studied Φ-Hiding assumption. For example, under this assumption, we show that RSA PKCS #1 v1.5 is secure against chosen-plaintext attacks for messages of length roughly $\frac{\log N}{4}$ bits, whereas the previous analysis, due to Lewko et al. [19], applies only to messages of length less than $\frac{\log N}{32}$.

In addition to providing new bounds, we also show that a key lemma of Lewko et al. [19] is incorrect. We prove a weaker version of the claim which is nonetheless sufficient for most, though not all, of their applications.

Our technical results can be viewed as showing that exponentiation in $\mathbb{Z}_N$ is a deterministic extractor for every source that is uniform on an arithmetic progression. Previous work showed this type of statement only on average over a large class of sources, or for much longer progressions (that is, sources with much more entropy).

*Computer Science and Engineering Department, Pennsylvania State University, University Park, PA 16802. {asmith,yxz169}@cse.psu.edu. A.S. and Y.Z. were supported by National Science Foundation awards #0747294 and #0941553. as well as a Google research award. Part of this work was done while A.S. was on sabbatical at Boston University's Hariri Institute for Computing.

1 Introduction

Cryptographic schemes based on the RSA trapdoor permutation [23] are ubiquitous in practice. Many of the schemes, are simple, natural and highly efficient. Unfortunately, their security is often understood only in the random oracle model [3], if at all.[1] When can the security of natural constructions be proven under well-defined and thoroughly studied assumptions? For example, consider the "simple embedding" RSA-based encryption scheme (of which RSA PKCS #1 v1.5, which is still in wide use, is a variant): given a plaintext x, encrypt it as $(x\|R)^e \bmod N$, where R is a random string of appropriate length and '$\|$' denotes string concatenation. Until recently [19], there was no proof of security for this scheme under a well-understood assumption. The security of this scheme under chosen plaintext attacks is closely related to another fundamental question, namely, whether many physical bits of RSA are simultaneously hardcore [2, 1].

Indistinguishability of RSA on Arithmetic Progressions. Both of these questions are related to the hardness of a basic computational problem, which we dub *RSA-AP*. Consider a game in which a distinguisher is first given an RSA public key (N, e) and a number K. The distinguisher then selects the description of an arithmetic progression (abbreviated "AP") $P = \{\sigma i + \tau \mid i = 0, \ldots, K-1\}$ of length K. Finally, the distinguisher gets a number $Y \in \mathbb{Z}_N$, and must guess whether Y was generated as $Y = X^e \bmod N$, where X is uniform in the AP P, or Y was drawn uniformly from $\mathbb{Z}_N$. We say RSA-AP is hard for length K (where K may depend on the security parameter) if no polynomial-time distinguisher can win this game with probability significantly better than it could by random guessing.

Hardness statements for the RSA-AP problem have important implications. For example, in the "simple embedding" scheme above, the input to the RSA permutation is $x\|R$, which is distributed uniformly over the AP $\{x2^\rho + i \mid i = 0 \ldots, 2^\rho - 1\}$ where ρ is the bit length of R. If RSA-AP is hard for length 2^ρ, then $(x\|R)^e \bmod N$ is indistinguishable from uniform for all messages x and so simple embedding is CPA secure.

In this paper, we show that RSA-AP is hard under well-studied assumptions, for much shorter lengths K than was previously known. From this, we draw conclusions about classic problems (the CPA security of PKCS #1 v1.5 and the simultaneous hardcoreness of many physical bits of RSA) that were either previously unknown, or for which previous proofs were incorrect.

Φ-Hiding, Lossiness and Regularity. The Φ-Hiding assumption, due to Cachin et al. [7], states that it is computationally hard to to distinguish standard RSA keys—that is, pairs (N, e) for which $\gcd(e, \Phi(N)) = 1$—from *lossy* keys (N, e) for which $e \mid \Phi(N)$. Under a lossy key, the map $x \mapsto x^e$ is not a permutation: if $N = pq$ where p, q are prime, e divides $p - 1$ and $\gcd(e, q - 1) = 1$, then $x \mapsto x^e$ is e-to-1 on $\mathbb{Z}_N^*$. We consider two variants for the lossy mode: one where p and q are chosen to have the same bit length, and one where their bit lengths differ by a specified difference θ (see Section 2).

The Φ-Hiding assumption has proven useful since under it, statements about *computational* indistinguishability in the real world (with regular keys) may be proven by showing the *statistical* indistinguishability of the corresponding distributions in the "lossy world" (where $e \mid \Phi(N)$) [18, 16, 19].

[1] There are many RSA-based constructions without random oracles, e.g., [5, 14, 15], but they are less efficient and not currently widely used.

Specifically, [19] showed that under Φ-Hiding, the hardness of RSA-AP for length K is implied by the approximate *regularity* of the map $x \mapsto x^e$ on arithmetic progressions when $e \mid \phi(N)$. Recall that a function is regular if it as the same number of preimages for each point in the image. For positive integers e, N and K, let $Reg(N, e, K, \ell_1)$ denote the maximum, over arithmetic progressions P of length K, of the statistical difference between $X^e \bmod N$, where $X \leftarrow_\$ P$, and a uniform e-th residue in $\mathbb{Z}_N$. That is,

$$Reg(N, e, K, \ell_1) \stackrel{def}{=} \max \left\{ SD(X^e \bmod N; U^e \bmod N) \;\middle|\; \begin{array}{l} \sigma \in \mathbb{Z}_N^*, \tau \in \mathbb{Z}_N, \\ X \leftarrow_\$ \{\sigma i + \tau \mid i = 0, \ldots, K-1\}, \\ U \leftarrow_\$ \mathbb{Z}_N \end{array} \right\}$$

Note that the maximum is taken over the choice of the AP parameters σ and τ. We can restrict our attention, w.l.o.g., to the case where $\sigma = 1$ (see Section 2); the maximum is thus really over the choice of τ.

Lewko et al. [19] observed that if $Reg(N, e, K, \ell_1)$ is negligible for the lossy keys (N, e), then Φ-Hiding implies that RSA-AP is hard for length K. Motivated by this, they studied the regularity of lossy exponentiation on arithmetic progressions. They claimed two types of bounds: average-case bounds, where the starting point τ of the AP is selected uniformly at random, and much weaker *worst-case* bounds, where τ is chosen adversarially based on the key (N, e).

1.1 Our Contributions

We provide new, *worst-case* bounds on the regularity of lossy exponentiation over $\mathbb{Z}_N$. These lead directly to new results on the hardness of RSA-AP, the CPA-security of simple padding-based encryption schemes, and the simultaneous hardcoreness of physical RSA bits. In addition, we provide a corrected version of the incorrect bound from [19] which allows us to recover some, though not all, of their claimed results.

Notice that in order to get any non-trivial regularity for exponentiation, we must have $K \geq N/e$, since there are at least N/e images. If the e-th powers of different elements were distributed uniformly and independently in $\mathbb{Z}_N$, then in fact we would expect statistical distance bounds of the form $\sqrt{\frac{N}{eK}}$. The e-th powers are of course not randomly scattered, yet we recover this type of distance bound under a few different conditions.

Our contributions can be broken into three categories:

Worst-case bounds (Section 3). We provide a new worst-case bound on the regularity of exponentiation for integers with an unbalanced factorization, where $q > p$. We show that

$$Reg(N, e, K, \ell_1) = O\left(\frac{p}{q} + \sqrt{\frac{N}{eK}}\right). \tag{1}$$

When q is much larger than p, our bound scales as $\sqrt{\frac{N}{eK}}$. This bound is much stronger than the analogous worst-case bound from Lewko et al. [19], which is $\tilde{O}\left(\sqrt{\frac{N}{eK}} \cdot \sqrt{\frac{N}{K}} \cdot \sqrt[8]{pe}\right)$ (where $\tilde{O}(\cdot)$ hides polylogarithmic factors in N).[2] In particular, we get much tighter bounds on the security of padding-based schemes than [19] (see "Applications", below).

[2]The bound of [19] relies on number-theoretic estimates of *Gauss sums*. Under the best known estimates [12], the bound has the form above. Even under the most optimistic number-theoretic conjecture on Gauss sums (the "MVW

Applying our new bounds requires one to assume a version of the Φ-Hiding assumption in which the "lossy" keys are generated in such a way that $q \gg p$ (roughly, $\log(q) \geq \log(p) + \lambda$ for security parameter λ). We dub this variant the *unbalanced* Φ-hiding assumption.

Perhaps surprisingly, the proof of our worst-case bounds in $\mathbb{Z}_N$ uses average-case bounds (for the smaller ring $\mathbb{Z}_p$), described next.

Average-case bounds (Section 4). We can remove the assumption that lossy keys have different-length factors if we settle for an average-case bound, where the average is taken over random translations of an arithmetic progression of a given length. We show that if X is uniform over an AP of length K, then

$$\mathop{\mathbb{E}}_{c \leftarrow\!\$\, \mathbb{Z}_N} \Big(SD\Big((c+X)^e \bmod N \;;\; U^e \bmod N \Big) \Big) = O\left(\sqrt{\tfrac{N}{eK}} + \tfrac{p+q}{N} \right),$$

where U is uniform in $\mathbb{Z}_N^*$. The expectation above can also be written as the distance between the pairs $(C, (C+X)^e \bmod N)$ and $(C, U^e \bmod N)$, where $C \leftarrow\!\$\, \mathbb{Z}_N$. This average-case bound is sufficient for our application to simultaneous hardcore bits.

This result was claimed in Lewko et al. [19] for *arbitrary* random variables X that are uniform over a set of size K. The claim is false in general (to see why, notice that exponentiation by e does not lose any information when working modulo q, and so $X \bmod q$ needs to be close to uniform in $\mathbb{Z}_q$). However, the techniques from our worst-case result can be used to prove the lemma for arithmetic progressions (and, more generally, for distributions X which are high in min-entropy and are distributed uniformly modulo q).

Applications (Section 5). Our bounds imply that, under Φ-Hiding, the RSA-AP problem is hard roughly as long as $K > \frac{N}{e}$. This, in turn, leads to new results on the security of RSA-based cryptographic constructions.

1. Simple encryption schemes that pad the message with a random string before exponentiating (including PKCS #1 v1.5) are semantically secure under unbalanced Φ-hiding as long as the random string is more than $\log(N) - \log(e)$ bits long (and hence the message is roughly $\log(e)$ bits). In contrast, the results of [19] only apply when the message has length at most $\frac{\log(e)}{16}$.[3] Known attacks on Φ-Hiding fail as long as $e \ll \sqrt{p}$ (see "Related Work", below). Thus, we can get security for messages of length up to $\frac{\log(N)}{4}$, as opposed to $\frac{\log(N)}{16}$. For example, when N is 8192 bits long, our analysis supports messages of 1735 bits with 80-bit security, as opposed to 128 bits [19].

2. Under Φ-hiding, the $\log(e)$ most (or least) significant input bits of RSA are simultaneously hardcore. This result follows from both types of bounds we prove (average- and worst-case). If we assume only that RSA is hard to invert, then the best known reductions show security only for a number of bits proportional to the security parameter (e.g., Akavia et al. [1]), which is at most $O(\sqrt[3]{\log N})$.

conjecture" [21]), the bounds of Lewko et al. [19] have the form $\tilde{O}(\sqrt{\frac{N}{eK}} \cdot \sqrt{\frac{N}{K}})$ and are consequently quite weak in the typical setting where $K \ll N$.

[3] Even under the MVW conjecture (see footnote 2), one gets security for messages of at most $\frac{\log(e)}{8}$ bits.

Lewko et al. [19] claimed a proof that *any* contiguous block of about $\log(e)$ physical bits of RSA is simultaneously hardcore. Our corrected version of their result applies on to the most or least significant bits, however. Proving security of other natural candidate hardcore functions remains an interesting open problem.

1.1.1 Techniques

The main idea behind our new worst-case bounds is to lift an average-case bound over the smaller ring $\mathbb{Z}_p$ to a worst-case bound on the larger ring $\mathbb{Z}_N$. First, note that we can exploit the product structure of $\mathbb{Z}_N \equiv \mathbb{Z}_q \times \mathbb{Z}_p$ to decompose the problem into mod p and mod q components. The "random translations" lemma of [19] *is* correct over $\mathbb{Z}_p$ (for p prime), even though it is false over $\mathbb{Z}_N$. The key observation is that, when the source X is drawn from a long arithmetic progression, the mod q component (which is close to uniform) acts as a random translation on the mod p component of X.

More specifically, let $V = [X \bmod q]$ denote the mod q component of X (drawn from an arithmetic progression of length much greater than q) and, for each value $v \in \mathbb{Z}_q$, let X_v denote the conditional distribution of X given $V = v$. Then

$$X_v \approx X_0 + v\,.$$

That is, X_v is statistically close to a translation of the shorter but sparser AP X_0 (namely, elements of the original AP which equal 0 modulo q). In the product ring $\mathbb{Z}_q \times \mathbb{Z}_p$, the random variable X is thus approximated by the pair

$$(\underbrace{V}_{\in \mathbb{Z}_q}, \underbrace{X_0 + V}_{\in \mathbb{Z}_p})\,.$$

Since V is essentially uniform in $\mathbb{Z}_q$, its reduction modulo p is also close to uniform in $\mathbb{Z}_p$ when $q \gg p$. This allows us to employ the random translations lemma in $\mathbb{Z}_p$ [19] to show that $X^e \bmod N$ is close to $U^e \bmod N$.

Discussion. Our worst-case bounds can be viewed as stating that multiplicative homomorphisms in $\mathbb{Z}_N$ (all of which correspond to exponentiation by a divisor of $\phi(N)$) are deterministic extractors for the class of sources that are uniform on arithmetic progressions of length roughly the number of images of the homomorphism. This is in line with the growing body of work in additive combinatorics that seeks to understand how additive and multiplicative structure interact. Interestingly, our proofs are closely tied to the product structure of $\mathbb{Z}_N$. The Gauss-sums-based results of Lewko et al. [19] remain the best known for analogous questions in $\mathbb{Z}_p$ when p is prime.

1.2 Related Work

Lossy trapdoor functions, defined by [22], are one of many concepts that allow us to deploy information-theoretic tools to analyze computationally-secure protocols. Lossiness is mostly used in the literature as a tool for designing new cryptographic systems. However, as mentioned above, Kiltz et al. [18] and Lewko et al. [19] showed that the concept also sheds light on existing constructions since, under the Φ-hiding assumption, the RSA permutation is lossy.

The Φ-hiding assumption predates those works considerably— introduced by Cachin et al. [7], it is the basis for a number of efficient protocols [7, 6, 10, 13]. Following [18], Kakvi and Kiltz [16]

showed that lossiness of RSA under ΦA is also useful to understand security of a classical RSA-based *signatures*. The best known cryptanalytic attack on Φ-hiding uses Coppersmith's technique for solving low-degree polynomials [8, 20] and applies when e is close to $p^{1/2}$ (the attack has a ratio of running time to success probability of at least $\sqrt{p}/e$, which implies that we should take $\log(e) \leq \frac{\log p}{2} - \lambda$ for security parameter λ). Other attacks [24] are for moduli of a special form that do not arise in the applications to RSA.

The security of the specific constructions we analyze has also been studied considerably. Bleichenbacher [4] gave *chosen-ciphertext* attacks on PKCS #1. Coron et al. [9] gave *chosen-plaintext* attacks for instantiations of PKCS #1 v1.5 encryption which pad the plaintext with a very short random string. In contrast, our security proofs require a large random string of at least $\frac{3}{4}\log N$ bits (though this is still shorter than the $\frac{7}{8}\log N$ random bits needed by the analysis of Lewko et al. [19]). Katz and Lindell [17, p. 363] mention that PKCS #1 v1.5 is believed to be CPA-secure for appropriate parameters, but no proof is known.

The "large hardcore bit conjecture" for RSA and the security of the simple embedding scheme are mentioned as important open problems by Goldreich [11]. Assuming that RSA is hard to invert implies only that λ bits are simultaneously hardcore, where 2^λ is the time needed to invert (see, e.g., [2, 1]). Prior progress was made by Steinfeld et al. [25], who showed that the $1/2 - 1/e - \epsilon - o(1)$ least significant bits of RSA are simultaneously hardcore under a computational problem related to the work of Coppersmith [8]. (This result does not apply directly to PKCS #1 v1.5 because the latter does not use the full RSA domain—some bits are fixed constants.)

2 Preliminaries

We denote by $SD(A; B)$ the statistical distance between the distributions of random variables A and B taking values in the same set. We write $A \approx_\epsilon B$ as shorthand for $SD(A; B) \leq \epsilon$. We consider adversaries that are restricted to probabilistic polynomial time (PPT), and let $negl(k)$ be a negligible function in k, that is, one that decreases faster than the reciprocal of any polynomial. We write $A \leftarrow\!\!\$\, B$ to indicate that the random variable A is generated by running (randomized) algorithm B using fresh random coins, if B is an algorithm, or that A is distributed uniformly in B, if B is a finite set.

Given an integer $I \in \mathbb{Z}^+$, we write $[I]$ for the set $\{0, 1, 2, \ldots, I-1\}$. Thus, an arithmetic progression ("AP") of length K can be written $P = \sigma[K] + \tau$ for some $\sigma, \tau \in \mathbb{Z}$.

Let Primes_t denote the uniform distribution of t-bit primes, and let $\mathsf{Primes}_t[\cdots]$ be shorthand the uniform distribution over t-bit primes that satisfy the condition in brackets. Let RSA_k denote the usual modulus generation algorithm for RSA which selects $p, q \leftarrow\!\!\$\, \mathsf{Primes}_{\frac{k}{2}}$ and outputs (N, p, q) where $N = pq$. Note that k is generally taken to be $\Omega(\lambda^3)$, where λ is the security parameter, so that known algorithms take 2^λ expected time to factor $N \leftarrow\!\!\$\, \mathsf{RSA}_k$.

The RSA-AP problem. The RSA-AP problem asks an attacker to distinguish $X^e \bmod N$ from $U^e \bmod N$, where $X \leftarrow\!\!\$\, P$ is drawn from an arithmetic progression and $U \leftarrow \mathbb{Z}_N$. We allow the attacker to choose the arithmetic progression based on the public key; this is necessary for applications to CPA security. We define $\mathsf{RSA\text{-}AP}(1^k, K)$ to be the assumption that the two following distributions are computationally indistinguishable, for any PPT attacker $\mathcal{A}$:

Experiment $\mathsf{RSA\text{-}AP}(1^k, K)$:	**Experiment** $\mathsf{RSA\text{-}Unif}(1^k, K)$:
$(N, p, q) \leftarrow\!\!\$\ \mathsf{RSA}_k$	$(N, p, q) \leftarrow\!\!\$\ \mathsf{RSA}_k$
$(\sigma, \tau) \leftarrow \mathcal{A}(N, e)$ where $\sigma \in \mathbb{Z}_N^*$ and $\tau \in \mathbb{Z}$	$(\sigma, \tau) \leftarrow \mathcal{A}(N, e)$
$X \leftarrow\!\!\$\ \{\sigma i + \tau : i = 0, \ldots K-1\}$	$U \leftarrow\!\!\$\ \mathbb{Z}_N$
Return (N, e, X)	Return (N, e, U)

Note that without loss of generality, we may always take $\sigma = 1$ in the above experiments, since given the key (N, e) and the element $X^e \bmod N$ where X is uniform in $P = \{\sigma i + \tau : i = 0, \ldots K-1\}$, one can compute $(\sigma^{-1}X)^e \bmod N$ where σ^{-1} is an inverse of σ modulo N. The element $\sigma^{-1}X$ is uniform in $P' = \left\{i + \sigma^{-1}\tau : i = 0, \ldots K-1\right\}$, while the element $\sigma^{-1}U$ will still be uniform in $\mathbb{Z}_N$. Hence, a distinguisher for inputs drawn from P can be used to construct a distinguisher for elements drawn from P', and vice-versa.

Φ-Hiding Assumption. Let θ be an even integer and $c \in (0, 1)$ be a constant. We define two alternate parameter generation algorithms for RSA keys:

Algorithm $RSA^{\text{inj}}_{c,\theta}(1^k)$:	**Algorithm** $RSA^{\text{loss}}_{c,\theta}(1^k)$
$e \leftarrow\!\!\$\ \mathsf{Primes}_{ck}$	$e \leftarrow\!\!\$\ \mathsf{Primes}_{ck}$
$(N, p, q) \leftarrow\!\!\$\ \mathsf{RSA}_k$	$p \leftarrow\!\!\$\ \mathsf{Primes}_{\frac{k}{2}-\frac{\theta}{2}}[p = 1 \bmod e]$
Return (N, e)	$q \leftarrow\!\!\$\ \mathsf{Primes}_{\frac{k}{2}+\frac{\theta}{2}}$
	Return (pq, e)

Definition 1 ((c, θ)-Φ-Hiding Assumption (ΦA))**.** Let θ, c be parameters that are functions of the modulus length k, where $\theta \in \mathbb{Z}^+$ is even and $c \in (0, 1)$. For any probabilistic polynomial-time distinguisher $\mathcal{D}$,

$$Adv^{\Phi A}_{c,\theta,\mathcal{D}}(k) = \left|\Pr[\mathcal{D}(RSA^{\text{inj}}_{c,\theta}(1^k)) = 1] - \Pr[\mathcal{D}(RSA^{\text{loss}}_{c,\theta}(1^k)) = 1]\right| \leq negl(k).$$

where $negl(k)$ is a negligible function in k.

As mentioned in the introduction, the regularity of lossy exponentiation on AP's of length K implies, under Φ-hiding, that $\mathsf{RSA\text{-}AP}$ is hard:

Observation 1. *Suppose that $Reg(N, e, K, \ell_1) \leq \epsilon$ for a $1 - \delta$ fraction of outputs of $RSA^{loss}_{c,\theta}(1^k)$. Then the advantage of an attacker $\mathcal{D}$ at distinguishing $\mathsf{RSA\text{-}AP}(1^k, K)$ from $\mathsf{RSA\text{-}Unif}(1^k)$ is at most $Adv^{\Phi A}_{c,\theta,\mathcal{D}}(k) + \epsilon + \delta$.*

Though the definitions above are stated in terms of asymptotic error, we state our main results directly in terms of a time-bounded distinguisher's advantage, to allow for a concrete security treatment.

3 Improved ℓ_1-Regularity Bounds for Arithmetic Progressions

Let $\mathcal{P} = \sigma[K] + \tau$ be an arithmetic progression where $K \in \mathbb{Z}^+$. In this section, we show that if X is uniformly distributed over an arithmetic progression, then $X^e \bmod N$ is statistically close to a uniformly random e-th residue in $\mathbb{Z}_N$.

Theorem 2. *Let $N = pq$ (p, q primes) and we assume $q > p$ and $\gcd(\sigma, N) = 1$. Let $\mathcal{P}$ be* AP *where $\mathcal{P} = \sigma[K] + \tau$ and assume that $K > q$. Let e be such that $e|p-1$ and $\gcd(e, q-1) = 1$. Then,*

$$SD(X^e \bmod N, U^e \bmod N) \leq \frac{3q}{K} + \frac{2p}{q-1} + \frac{2}{p-1} + \sqrt{\frac{N}{eK}}$$

where $X \leftarrow_\$ \mathcal{P}$ and $U \leftarrow_\$ \mathbb{Z}_N^*$.

Recall, from Section 2, that it suffices to prove the Theorem for $\sigma = 1$. The main idea behind the proof is as follows. For any $v \in \mathbb{Z}_q$ and a set $\mathcal{P} \subset \mathbb{Z}_N$, we define $\mathcal{P}_v = \{x \in \mathcal{P} | x \bmod q = v\}$. First, we observe that $SD(X^e \bmod N, U^e \bmod N) \approx \mathbb{E}_{v \in \mathbb{Z}_q^*}(SD(X_v^e \bmod p, U_p^e \bmod p))$ (Lemma 3) where $U_p \leftarrow_\$ \mathbb{Z}_p^*$ and for any $v \in \mathbb{Z}_q^*$, $X_v \leftarrow_\$ \mathcal{P}_v$. Second, we show that $\mathcal{P}_v$ is almost identical to $\mathcal{P}_0 + v$ (that is the set $\mathcal{P}_0$ shifted by $v \in \mathbb{Z}_q$) (Lemma 4). Therefore, we can replace $\mathbb{E}_{v \leftarrow_\$ \mathbb{Z}_q^*}(SD(X_v^e \bmod p, U_p^e \bmod p))$ with $\mathbb{E}_{v \leftarrow_\$ \mathbb{Z}_q^*}(SD((Y + v')^e \bmod p, U_p^e \bmod p))$ where $Y \leftarrow_\$ \overline{\mathcal{P}}$. The last term can be bounded via hybrid arguments and a similar technique to [19, Lemma 3] (our Lemma 6).

In order to prove this theorem, we need the following lemmas (whose proof will be given at the end of this section):

Lemma 3. *Let $N = pq$ (p, q primes). Let $\mathcal{P}$ be an* AP *where $\mathcal{P} = [K] + \tau$ and assume that $K > q$. Let e be such that $e|p-1$ and $\gcd(e, q-1) = 1$. Then,*

$$SD(X^e \bmod N, U^e \bmod N) \leq \frac{q}{K} + \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*}(SD(X_v^e \bmod p, U_p^e \bmod p))$$

where $X \leftarrow_\$ \mathcal{P}$, $U \leftarrow_\$ \mathbb{Z}_N^*$, $U_p \leftarrow_\$ \mathbb{Z}_p^*$ and for any $v \in \mathbb{Z}_q^*$, $X_v \leftarrow_\$ \mathcal{P}_v$.

Lemma 4. *Let $N = pq$ (p, q primes). Let $\mathcal{P}$ be an* AP *where $\mathcal{P} = [K] + \tau$. For any $v \in \mathbb{Z}_q^*$, $|\mathcal{P}_v \triangle (\mathcal{P}_0 + v)| \leq 2$ where $\triangle$ denotes symmetric difference.*

Lemma 5. *Let $N = pq$ (p, q primes) and assume $q > p$. Let e be such that $e|p-1$ and $\gcd(e, q-1) = 1$. Let $\overline{\mathcal{K}} \subset \mathbb{Z}_N$ be an arbitrary subset (not necessarily an* AP*):*

$$\begin{aligned} &SD((C \bmod p, (C+R)^e \bmod p), (C \bmod p, U_p^e \bmod p)) \\ &\leq SD((V_p, (V_p+R)^e \bmod p), (V_p, U_p^e \bmod p)) + \frac{2p}{q-1}. \end{aligned}$$

where $C \leftarrow_\$ \mathbb{Z}_q^*, V_p, U_p \leftarrow_\$ \mathbb{Z}_p^*$ and $R \leftarrow_\$ \overline{\mathcal{K}}$.

Notice that in this lemma, the random variable C is chosen from $\mathbb{Z}_q^*$ but always appears reduced modulo p.

Roughly speaking, Lemma 5 says that if $[I]$ ($I \in \mathbb{Z}^+$; e.g., $I = q-1$) is large enough ($I > p$), we can replace $Q \bmod p$ with V_p, where $Q \leftarrow_\$ [I]$ and $V_p \leftarrow_\$ \mathbb{Z}_p^*$. Then, we can apply the random translations lemma [19] over $\mathbb{Z}_p^*$ to show Lemma 6.

We should point out that the mistake in the proof of [19] does not apply to Lemma 6. Specifically, the mistake in [19] is due to the fact that $\omega - 1$ may not be invertible in $\mathbb{Z}_N$ where $N = pq$, $\omega^e = 1 \bmod N$ and $\omega \neq 1$ (refer Section 4 for more detailed explanation). However, $\omega - 1$ is invertible in $\mathbb{Z}_p$, (since p is prime) which is the ring used in Lemma 6. Specifically, we apply the following corrected version of [19, Lemma 3]:

Lemma 6 (Random Translations Lemma, adapted from [19])**.** *Let $N = pq$ (p, q primes). Let $V_p, U_p \leftarrow_\$ \mathbb{Z}_p^*$. Let $R \leftarrow_\$ \overline{\mathcal{K}}$ where $\overline{\mathcal{K}} \subset \mathbb{Z}_N$ and $|\overline{\mathcal{K}}| = \overline{K}$.*

$$SD((V_p, (V_p + R)^e \bmod p), (V_p, U_p^e \bmod p)) \leq \frac{2}{p-1} + \sqrt{\frac{p-1}{e\overline{K}}}.$$

The proof of Lemma 6 is given in Appendix A. We can now prove our main result, Theorem 2:

Proof of Theorem 2. Let $X \leftarrow_\$ \mathcal{P}$, $U \leftarrow_\$ \mathbb{Z}_N^*$, $U_p \leftarrow_\$ \mathbb{Z}_p^*$. For any $v \in \mathbb{Z}_q$, let $X_v \leftarrow_\$ \mathcal{P}_v$ (recall $\mathcal{P}_v$ is a set $\{x \in \mathcal{P} | x \bmod q = v\}$). By Lemma 3, we have:

$$SD(X^e \bmod N, U^e \bmod N) \leq \frac{q}{K} + \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} (SD(X_v^e \bmod p, U_p^e \bmod p)).$$

Let $Y \leftarrow_\$ \mathcal{P}_0$. By the triangle inequality:

$$\mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} SD(X_v^e \bmod p, U_p^e \bmod p) \leq \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} (SD(X_v, Y + v) + SD((Y + v)^e \bmod p, U_p^e \bmod p)).$$

Note that $SD(A^e \bmod p, B^e \bmod p) \leq SD(A, B)$ for any A and B. By Lemma 4, for every $v \in \mathbb{Z}_q^*$, we have $|\mathcal{P}_v \triangle (\mathcal{P}_0 + v)| \leq 2$. Therefore, we have $SD(X_v, Y + v) = \frac{|\mathcal{P}_v \triangle (\mathcal{P}_0 + v)|}{|\mathcal{P}_0|} \leq \frac{2}{|\mathcal{P}_0|} \leq \frac{2q}{K}$. Then,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} \big(SD(X_v, Y + v) + SD((Y + v)^e \bmod p, U_p^e \bmod p)\big) \\
&\leq \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} SD(X_v, Y + v) + \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} SD\left((Y + v)^e \bmod p, U_p^e \bmod p\right) \\
&\leq \frac{2q}{K} + \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} SD\left((Y + v)^e \bmod p, U_p^e \bmod p\right).
\end{aligned}$$

First, note that only the reduced value of $v \bmod p$ affects the statistical distance $SD((Y + v)^e \bmod p, U_p^e \bmod p)$. so the expression above can be rewritten as:

$$\mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*} SD(((Y + v)^e \bmod p, U_p^e \bmod p)) = \mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*; w \leftarrow_\$ v \bmod p} SD\left((Y + w)^e \bmod p, U_p^e \bmod p\right).$$

Let $U_q \leftarrow_\$ \mathbb{Z}_q^*$. The expectation above can be written as the distance between two pairs:

$$\mathop{\mathbb{E}}_{v \leftarrow_\$ \mathbb{Z}_q^*; w \leftarrow_\$ v \bmod p} SD((Y + w)^e \bmod p, U_p^e \bmod p) = SD\left(U_q \bmod p, (Y + U_q)^e \bmod p, (U_q \bmod p, U_p^e \bmod p)\right).$$

By Lemma 3 and 4, we have $SD\left(U_q \bmod p, (R + U_q)^e \bmod p, (U_q \bmod p, U_p^e \bmod p)\right) < \frac{2p}{q-1} + \frac{2}{p-1} + \sqrt{\frac{p-1}{e|\overline{\mathcal{K}}|}}$ where $\overline{\mathcal{K}} \subset \mathbb{Z}_N$ and $R \leftarrow_\$ \overline{\mathcal{K}}$. We apply the inequality with $\overline{\mathcal{K}} = \mathcal{P}_0$:

$$\begin{aligned}
&SD((U_q \bmod p, (Y + U_q)^e \bmod p), (U_q \bmod p, U_p^e \bmod p)) \\
&\leq \frac{2p}{q-1} + \frac{2}{p-1} + \sqrt{\frac{p-1}{e|\mathcal{P}_0|}} \leq \frac{2p}{q-1} + \frac{2}{p-1} + \sqrt{\frac{N}{eK}}.
\end{aligned}$$

since $|\mathcal{P}_0| = \{\lfloor \frac{K}{q} \rfloor, \lceil \frac{K}{q} \rceil\}$. □

3.1 Proofs of Lemmas

We now prove the technical lemmas from previous section.

Proof of Lemma 3. The proof is done via hybrid arguments. By the Chinese Remainder Theorem, the mapping $a \mapsto (a \bmod p, a \bmod q)$ is an isomorphism from $\mathbb{Z}_N \to \mathbb{Z}_p \times \mathbb{Z}_q$. Therefore, we can rewrite $SD(X^e \bmod N, U^e \bmod N)$ as $SD((X^e \bmod p, X^e \bmod q), (U_p^e \bmod p, U_q^e \bmod q))$ where $U \leftarrow_\$ \mathbb{Z}_N^*$, $U_p \leftarrow_\$ \mathbb{Z}_p^*$ and $U_q \leftarrow_\$ \mathbb{Z}_q^*$. Furthermore, as $\gcd(e, q-1) = 1$, $a \to a^e \bmod q$ is a 1-to-1 mapping over $\mathbb{Z}_q^*$. Therefore,

$$\begin{aligned} &SD((X^e \bmod p, X^e \bmod q), (U_p^e \bmod p, U_q^e \bmod q)) \\ &= SD((X^e \bmod p, X \bmod q), (U_p^e \bmod p, U_q \bmod q). \end{aligned}$$

Now, we define $T_0 = (X \bmod q, X^e \bmod p)$, $T_1 = (U_q, X_{U_q}^e \bmod p)$ and $T_2 = (U_q, U_p^e \bmod p)$ where X_{U_q} is the random variable that chooses $v \leftarrow_\$ \mathbb{Z}_q^*$ and then $X_{U_q} \leftarrow_\$ \mathcal{P}_v$. By the triangle inequality (hybrid arguments),

$$SD(T_0, T_2) \le SD(T_0, T_1) + SD(T_1, T_2)$$

where we have $SD(T_1, T_2) = \mathbb{E}_{v \in \mathbb{Z}_q^*} SD((X_v^e \bmod p, U_p^e \bmod p)$.

Now, we bound $SD(T_0, T_1)$. Define $T_0' = (W \bmod q, X_{W \bmod q}^e \bmod p)$ where $W \leftarrow_\$ [K]$ (recall that $|\mathcal{P}| = K$). We claim that $SD(T_0, T_1) = SD(T_0', T_1)$. Specifically,

$$\begin{aligned} SD(T_0, T_1) &= \frac{1}{2} \sum_{a \in \mathbb{Z}_q^*} \left| \Pr_{(\ell+\tau) \leftarrow_\$ \mathcal{K}}[\ell + \tau \bmod q = a] - \Pr_{x \leftarrow_\$ \mathbb{Z}_q^*}[x \bmod q = a] \right| \\ &= \frac{1}{2} \sum_{a \in \mathbb{Z}_q^*} \left| \Pr_{\ell \leftarrow_\$ [K]}[\ell \bmod q = (a - \tau) \bmod q] - \Pr_{x \leftarrow_\$ \mathbb{Z}_q^*}[x \bmod q = a] \right| \\ &= \frac{1}{2} \sum_{a \in \mathbb{Z}_q^*} \left| \Pr_{\ell \leftarrow_\$ [K]}[\ell \bmod q = (a - \tau) \bmod q] - \Pr_{x \leftarrow_\$ \mathbb{Z}_q^*}[x \bmod q = (a - \tau) \bmod q] \right| \\ &= SD(T_0', T_1). \end{aligned}$$

Now, we bound $SD(T_0', T_1)$:

$$\begin{aligned} SD(T_0', T_1) &= SD((W \bmod q, X_{W \bmod q}^e \bmod p), (U_q, X_{U_q}^e \bmod p)) \\ &\le SD(W \bmod q, U_q). \end{aligned}$$

Let $r = K \bmod q$. Then,

$$\begin{aligned} SD(W \bmod q, U_q) &= \frac{1}{2} \sum_{a \in \mathbb{Z}_q^*} \left| \Pr_{x \leftarrow_\$ [K]}[x \bmod q = a] - \Pr_{x \leftarrow_\$ \mathbb{Z}_q^*}[x = a] \right| \\ &= r\left(\frac{(K-r)/q + 1}{K} - \frac{1}{q-1}\right). \end{aligned}$$

Note that $\frac{(K-r)/q+1}{K} \leq (1+\frac{q-r}{K})\frac{1}{q-1}$ and we have:

$$r(\frac{(K-r)/q+1}{K} - \frac{1}{q-1}) \leq \frac{r}{(q-1)}\frac{q-r}{K} \leq \frac{q}{K}$$

as $0 \leq r \leq q-1$. To conclude,

$$\begin{aligned} SD(T_0, T_2) &\leq SD(T_0', T_1) + SD(T_1, T_2) \\ &\leq \frac{q}{K} + \mathop{\mathbb{E}}_{v \leftarrow\$ \mathbb{Z}_q^*} SD((X_v^e \bmod p, U_p^e \bmod p)). \end{aligned}$$

□

Proof of Lemma 4. Let $u \in \mathbb{Z}_q$, we have

$$\begin{aligned} \mathcal{P}_u &= \{x \in \mathcal{P} | x \bmod q = u\} \\ &= \{\ell + \tau | \ell \leq K \wedge \ell = u - \tau \bmod q\} \\ &= \{(u-\tau) \bmod q + qk + \tau | 0 \leq k \leq \frac{K - (u-\tau) \bmod q}{q}\}. \end{aligned}$$

Specifically, we have $\mathcal{P}_0 = \{qk - \tau \bmod q + \tau | 0 \leq k \leq \frac{K+\tau \bmod q}{q}\}$. Recall that $v < q$ ($v \in \mathbb{Z}_q^*$), we have:

$$\mathcal{P}_v = \begin{cases} \{qk - \tau \bmod q + \tau + q + v | 0 \leq k \leq \frac{K-v+\tau \bmod q)}{q} - 1\} & v < \tau \bmod q; \\ \{qk - \tau \bmod q + \tau + v | 0 \leq k \leq \frac{K-v+\tau \bmod q)}{q}\} & \text{otherwise.} \end{cases}$$

Therefore, for any $v \in \mathbb{Z}_q^*$, $|\mathcal{P}_v \triangle (\mathcal{P}_0 + v)| \leq 2$ where $\triangle$ denotes symmetric difference. □

Proof of Lemma 5. The proof is done via hybrid arguments. Let $T_0 = (C \bmod p, (C \bmod p + R)^e \bmod p)$, $T_1 = (V_p, (V_p + R)^e \bmod p)$ and $T_2 = (V_p, U_p^e \bmod p)$ and $T_3 = (C \bmod p, U_p^e \bmod p)$. Then,

$$SD(T_0, T_3) \leq SD(T_0, T_1) + SD(T_1, T_2) + SD(T_2, T_3).$$

Via the similar technique (to show $SD(W \bmod q, U_q)$) in Lemma 3, we have:

$$\begin{aligned} SD(T_0, T_1) = SD(T_2, T_3) &= SD(C \bmod p, U_p) \\ &\leq \frac{p}{|C|} = \frac{p}{q-1}. \end{aligned}$$

□

4 Average-case Bounds over Random Translations

In this section, we point out a mistake in the proof of Lemma 4 from [19]. We give a counter example to the lemma, explain the error in the proof and prove a corrected version of the lemma which still implies the main conclusions from [19]. First, we restate their lemma:

Incorrect Claim 1 (Lemma 4 [19])**.** *Let $N = pq$ and e be such that $e|p-1$ and $gcd(e, q-1) = 1$. Let $\mathcal{K} \subset \mathbb{Z}_N$ such that $|\mathcal{K}| \geq \frac{4N}{e\alpha^2}$ for some $\alpha \geq \frac{4(p+q-1)}{N}$. Then,*

$$SD((C, (C+X)^e \bmod N), (C, U^e \bmod N) \leq \alpha$$

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

4.1 Counterexample to Lemma 4 in LOS

The problem with this lemma, as stated, is that raising numbers to the e-th power is a permutation in $\mathbb{Z}_q$, and so exponentiation does not erase any information (statistically) about the value of the input mod q. (It may be that information is lost computationally when p, q are secret, but the claim is about statistical distance.)

Adding a publicly available random offset does not help, since the composition of translation and exponentiation is still a permutation of $\mathbb{Z}_q$. Hence, if $X \bmod q$ is not close to uniform, then $(C, (C+X)^e \bmod q)$ is not close to uniform in $\mathbb{Z}_N \times \mathbb{Z}_q$, and so $(C, (C+X)^e \bmod N)$ is not close to uniform in $\mathbb{Z}_N^2$.

To get a counterexample to the claimed lemma, let $\mathcal{K} = \left\{ x \in \mathbb{Z}_N : x \bmod q \in \{0, ..., \frac{q-1}{2}\} \right\}$ (the subset of $\mathbb{Z}_N$ with mod q component less than $q/2$). $\mathcal{K}$ is very large (size about $N/2$) but the pair $C, (X+C)^e \bmod q$ will never be close to uniform when $X \leftarrow\!\!\$\, \mathcal{K}$.

The above attack was motivated by the discovery of a mistake in the proof of Lemma 4 from [19]. Specifically, the authors analyze the probability that $(C+X)^e = (C+Y)^e$ by decomposing the event into events of the form $(C+X) = \omega(C+Y)$ where ω is an e-th root of unity. The problem arises because

$$\Pr[(C+X) = \omega(C+Y)] \neq \Pr[C = (\omega - 1)^{-1}(\omega Y - X)]$$

since $\omega - 1$ is not invertible in $\mathbb{Z}_N^*$ (it is 0 mod q).

4.2 Corrected Translation Lemma

It turns out that distinguishability mod q is the only obstacle to the random translation lemma. We obtain the following corrected version:

Lemma 7. *Let $N = pq$ and e be such that $e|p-1$ and $\gcd(e, q-1) = 1$. Let $\mathcal{K} \subset \mathbb{Z}_N$ be an arithmetic progression. Specifically, let $\mathcal{K} = \sigma[K] + \tau$ with $K > q$. Then,*

$$\begin{aligned} SD\big((C, (C+X)^e \bmod N), (C, U^e \bmod N)\big) &\leq \frac{1}{p} + \frac{2}{p-1} + \sqrt{\frac{N}{eK}} + SD(X \bmod q, U \bmod q) \\ &\leq \frac{3}{p-1} + \sqrt{\frac{N}{eK}} + \frac{q}{K}. \end{aligned}$$

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

Proof. Applying the same idea in Lemma 3, let $U_p \leftarrow\!\!\$\, \mathbb{Z}_p, U_q \leftarrow\!\!\$\, \mathbb{Z}_q$, we have:

$$\begin{aligned} &SD\big((C, (C+X)^e \bmod N), (C, U^e \bmod N)\big) \\ &= \mathop{\mathbb{E}}_{c \leftarrow\!\!\$\, \mathbb{Z}_N} \left(SD((c+X)^e \bmod N, U^e \bmod N)\right) \\ &= \mathop{\mathbb{E}}_{c \leftarrow\!\!\$\, \mathbb{Z}_N} \big(SD(((c+X)^e \bmod p, (c+X) \bmod q), (U_p^e \bmod p, U_q))\big). \end{aligned}$$

Notice that the mod$q$ components are not raised to the e-th power. This is because exponentiation is a permutation of $\mathbb{Z}_q^*$ as $\gcd(e, q-1) = 1$. For any $c \in \mathbb{Z}_N$, let $T_0(c) = ((c+X)^e \bmod p, (c+X) \bmod q), T_1(c) = ((c+X)^e_{U_q} \bmod p, U_q), T_2 = (U_p^e \bmod p, U_q)$. Then, we can rewrite

$\mathbb{E}_{c \leftarrow\$ \mathbb{Z}_N} \left(SD(((c+X)^e \bmod p, (c+X) \bmod q), (U_p^e \bmod p, U_q)) \right)$ as $\mathbb{E}_{c \leftarrow\$ \mathbb{Z}_N} SD\left(T_0(c), T_2\right)$. By the triangle inequality, we have:

$$\mathop{\mathbb{E}}_{c \leftarrow\$ \mathbb{Z}_N} SD(T_0(c), T_2) \leq \mathop{\mathbb{E}}_{c \leftarrow\$ \mathbb{Z}_N} (SD(T_0(c), T_1(c)) + SD(T_1(c), T_2)).$$

For each $c \in \mathbb{Z}_N$:

$$\begin{aligned} &SD(T_0(c), T_1) = SD\Big(((c+X)^e \bmod p, (c+X) \bmod q), ((c+X)^e_{U_q} \bmod p, U_q)\Big) \\ &\leq SD((c+X) \bmod q, U_q) \leq SD(X \bmod q, U_q). \end{aligned}$$

The last equality holds because translation by c is a permutation of $\mathbb{Z}_q$. We have:

$$\begin{aligned} &SD(T_1(c), T_2) = SD\Big(((c+X)^e_{U_q} \bmod p, U_q), (U_p^e \bmod p, U_q)\Big) \\ &= \mathop{\mathbb{E}}_{v \leftarrow\$ \mathbb{Z}_q} SD((X+c)^e_v \bmod p, U_p^e \bmod p). \end{aligned}$$

Recall that for any $v \in \mathbb{Z}_q$, $(c+X)_v$ denotes the random variable $c+X$ conditioned on the event that $c+X \bmod q = v$. To sum up,

$$\begin{aligned} &\mathop{\mathbb{E}}_{c \leftarrow\$ \mathbb{Z}_N} ((c+X)^e \bmod N, U^e \bmod N) \\ &\leq SD(X \bmod q, U_q) + \mathop{\mathbb{E}}_{v \leftarrow\$ \mathbb{Z}_q} \mathop{\mathbb{E}}_{c \leftarrow\$ \mathbb{Z}_N} SD\left((X+c)^e_v \bmod p, U_p^e \bmod p\right). \end{aligned}$$

Note that only the value of $c \bmod p$ affects $SD((X+c)^e_v \bmod p, U_p^e \bmod p)$. We can replace $c \leftarrow\$ \mathbb{Z}_N$ with $V_p \leftarrow\$ \mathbb{Z}_p^*$. Specifically, let **BAD** be the event that $\gcd(c,p) \neq 1$. As $c \leftarrow\$ \mathbb{Z}_N$, we have $\Pr[\mathbf{BAD}] = \Pr_{c \leftarrow\$ \mathbb{Z}_N}[\gcd(c,p) \neq 1] = \frac{1}{p}$. Therefore, for any $v \in \mathbb{Z}_q$,

$$\begin{aligned} &\mathop{\mathbb{E}}_{c \leftarrow\$ \mathbb{Z}_N} SD((X+c)^e_v \bmod p, U_p^e \bmod p) \\ &\leq \Pr[\mathbf{BAD}] \cdot 1 + 1 \cdot \mathop{\mathbb{E}}_{c \leftarrow\$ \mathbb{Z}_p^*} SD((X+c)^e_v \bmod p, U_p^e \bmod p) \\ &\leq \frac{1}{p} + \mathop{\mathbb{E}}_{V_p \leftarrow\$ \mathbb{Z}_p^*} SD((X+V_p)^e_v \bmod p, U_p^e \bmod p) \end{aligned}$$

as $\Pr[\mathbf{BAD}] < 1$ and statistical distance $SD(\cdot,\cdot) < 1$.

By Lemma 6, we have $\mathbb{E}_{V_p \leftarrow\$ \mathbb{Z}_p^*} \left((X+V_p)^e_v \bmod p, U_p^e \bmod p\right) \leq \frac{2}{p-1} + \sqrt{\frac{N}{eK}}$. Thus,

$$\begin{aligned} &SD_{C \leftarrow\$ \mathbb{Z}_N}((C, (C+X)^e \bmod N), (C, U^e \bmod N)) \\ &\leq \frac{1}{p} + \frac{2}{p-1} + \sqrt{\frac{N}{eK}} + SD(X \bmod q, U_q) \leq \frac{1}{p} + \frac{2}{p-1} + \sqrt{\frac{N}{eK}} + \frac{q}{K}. \end{aligned}$$

□

5 Applications

In this section, we apply the above results to understanding the IND-CPA security of PKCS #1 v1.5 and to showing that the most/least $\log e - 3\log\frac{1}{\epsilon} + o(1)$ significant RSA bits are simultaneously hardcore. To illustrate our results, we show that our bounds imply improvements to the concrete security parameters from Lewko et al. [19].

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

Below, a_{16} denotes the 16-bit binary representation of a two-symbol hexadecimal number $a \in \{00, ..., FF\}$. Let $PKCS(x; r) = x||00_{16}||r$. The ciphertext for message x under PKCS #1 v1.5 [4] is then $(00_{16}||02_{16}||PKCS(x; r))^e \bmod N$, where r is chosen uniformly random from $\{0,1\}^\rho$.

Theorem 8 (CPA security of PKCS #1 v1.5)**.** *Let λ be the security parameter, $k = k(\lambda) \in \mathbb{Z}^+$ and $\epsilon(\lambda), c(\lambda) > 0$. Suppose ΦA holds for c and $\theta \geq 4 + \log \frac{1}{\epsilon}$. Let Π_{PKCS} be the PKCS #1 v1.5 encryption scheme. Assume that $\rho \geq \log N - \log e + 2\log(1/\epsilon) + 4$ and $\theta \geq 4 + \log \frac{1}{\epsilon}$. Then for any IND-CPA adversary $\mathcal{A}$ against Π_{PKCS}, there exists a distinguisher $\mathcal{D}$ for Φ-Hiding with $time(\mathcal{D}) \leq time(\mathcal{A}) + O(k^3)$ such that for all $\lambda \in \mathbb{N}$:*

$$Adv^{ind-cpa}_{\Pi_{PKCS},\mathcal{A}}(\lambda) \leq Adv^{\Phi A}_{c,\theta,\mathcal{D}}(\lambda) + \epsilon(\lambda).$$

Proof. Define $\mathbf{Game}_0$ be the original IND-CPA security game with the adversary $\mathcal{A}$. Let $\mathbf{Game}_1$ be identical to $\mathbf{Game}_0$ except that (N, e) is generated via lossy RSA key generation (Section 2, Φ-Hiding Assumption), such that $e|p-1$ and $gcd(e, q-1) = 1$. $\mathbf{Game}_2$ is identical to $\mathbf{Game}_1$ except that the challenge ciphertext $c^* = (00_{16}||02_{16}||PKCS(x^*, r^*))^e \bmod N$ is replaced with $U^e \bmod N$ where $U \leftarrow\!\!\$\, \mathbb{Z}^*_N$.

An adversary who performs differently in $\mathbf{Game}_0$ and $\mathbf{Game}_1$ can be used to attack the ΦA assumption; the increase in running time is the time it takes to generate a challenge ciphertext, which is at most $O(k^3)$. The difference between $\mathbf{Game}_1$ and $\mathbf{Game}_2$ is $SD((00_{16}||02_{16}||PKCS(x^*, r^*))^e \bmod N, (\mathbb{Z}^*_N)^e \bmod N)$ (information theoretically) where x^* is the challenge plaintext and r^* is the encryption randomness. Specifically, given the challenge plaintext x^* that may depend on $pk = (N, e)$, $00_{16}||02_{16}||PKCS(x^*, \cdot) = \{r + x^*2^{\rho+8} + 2^{\rho+8+|x|} | r \in \{0,1\}^\rho\}$ is an arithmetic progression with length 2^ρ. By Theorem 2,

$$\begin{aligned} SD(0002_{16}||PKCS(x^*, r^*)^e \bmod N, (\mathbb{Z}^*_N)^e \bmod N) &\leq \frac{1}{p-1} + \frac{2p}{q-1} + \frac{3q}{2^{\rho+1}} + \sqrt{\frac{N}{e2^\rho}} \\ &\leq 2\left(\frac{2p}{q-1} + \sqrt{\frac{N}{e2^\rho}}\right) < \epsilon \end{aligned}$$

where we have $\frac{2p}{q-1} < \frac{\epsilon}{4}$ when $\theta \geq 4 + \log \frac{1}{\epsilon}$, and $\sqrt{\frac{N}{e2^\rho}} < \frac{\epsilon}{4}$ when $\rho \geq \log N - \log e - 2\log \epsilon + 4$. Note that the advantage of $\mathcal{A}$ in $\mathbf{Game}_2$ is 0. □

Achievable Parameters. To get a sense of the parameters for which our analysis applies, recall the best known attack on Φ-Hiding (using Coppersmith's algorithm) has a tradeoff of time to success probability of at least 2^λ when $p < q$ and $\log(e) = \frac{\log(p)}{2} - \lambda$. We therefore select this value of e (that is, $e = \sqrt{p}/2^\lambda$) for security parameter λ.

For a message of length m, PKCS #1 v1.5 uses a random string of length $\rho = \log N - m - 48$ (since six bytes of the padded string are fixed constants). To apply Theorem 8, we need two conditions. First, we need $\rho \geq \log N - \log e + 2\log(1/\epsilon) + 4$; for this, it suffices have a message of length $m \leq \log(e) - 2\log(1/\epsilon) - 52$. Second, we need $\theta = \log q - \log p \geq 4 + \log(1/\epsilon)$. Setting p to have length $\log(p) = \frac{\log(N)}{2} - \frac{\log(1/\epsilon)+4}{2}$ satisfies this condition.

[4] RFC2313, http://tools.ietf.org/html/rfc2313

Using the value of e based on Coppersmith's attack, and setting $\epsilon = 2^{-\lambda}$ in Theorem 8, we get CPA security for messages of length up to

$$m = \tfrac{1}{4}\log N - \tfrac{13}{4}\lambda - 53\,. \tag{2}$$

with security parameter λ.

In contrast, the analysis of Lewko et al. [19] proves security for messages of length only $m = \frac{\log N}{16} - \Theta(\lambda)$. Even under the most optimistic number-theoretic conjecture (the MVW conjecture on Gauss sums), their analysis applies to messages of length only $m = \frac{\log N}{8} - \Theta(\lambda)$. [5] Their proof methodology cannot go beyond that bound. Our results therefore present a significant improvement over the previous work.

Concrete Parameters: Take the modulus length $k = \log N = 8192$ as an example. We will aim for $\lambda = 80$-bit security. We get CPA security for messages of length up to

$$m = \frac{\log N}{4} - \frac{13}{4}\lambda - 53 = 1735 \text{ (bits)}.$$

This is improves over the 128 bit messages supported by the analysis of [19] by a factor of 13. (That said, we do not claim to offer guidance for setting parameters in practice, since our results require an exponent e much larger than the ones generally employed.)

5.2 (Most/Least Significant) Simultaneously Hardcore Bits for RSA

Let λ be the security parameter and let $k = \log N$ be the modulus length. For $1 \le i < j \le k$, we want to show that the two distributions $(N, e, x^e \bmod N, x[i,j])$ and $(N, e, x^e \bmod N, r)$ are computationally indistinguishable, where $x \leftarrow\!\!\$\, \mathbb{Z}_N^*$, $r \leftarrow\!\!\$\, \{0,1\}^{j-i-1}$, and $x[i:j]$ denotes bits i through j of the binary representation of x.

In this section, we apply Theorem 2 to show the most and least $\log e - O(\log \frac{1}{\epsilon})$ significant bits of RSA functions are simultaneously hardcore (Theorem 9). We should note that we can apply the corrected random translations lemma (our Lemma 7) to this problem, which yields an essentially identical result. For brevity, we omit its proof.

Theorem 9. *Let λ be the security parameter, $k = k(\lambda) \in \mathbb{Z}^+$ and $\epsilon(\lambda), c(\lambda) > 0$. Suppose ΦA holds for c and $\theta > 4 + \log \frac{1}{\epsilon}$. Then, the most (or least) $\log e - 2\log \frac{1}{\epsilon} - 2$ significant bits of RSA are simultaneously hardcore. Specifically, for any distinguisher $\mathcal{D}$, there exists a distinguisher $\overline{\mathcal{D}}$ running in time time$(\mathcal{D}) + O(k^3)$ such that*

$$\left|\Pr[\mathcal{D}(N, e, x^e \bmod N, x[i:j]) = 1] - \Pr[\mathcal{D}(N, e, x^e \bmod N, r[i:j]) = 1]\right| \le Adv^{\Phi A}_{c,\theta,\overline{\mathcal{D}}}(\lambda) + 2\epsilon.$$

where $r \leftarrow\!\!\$\, \mathbb{Z}_N$, $|j - i| \le \log e - 2\log \frac{1}{\epsilon} - 2$ and either $i = 1$ or $j = k$. Furthermore, the distribution of $r[i;j]$ is 2^{k-j}-far from uniform on $\{0,1\}^{j-i+1}$.

It's important to note that the theorem is stated in terms of the distinguishability between bits i through j of the RSA input, and bits i through j of a random element r of $\mathbb{Z}_N$. The string $r[i:j]$ is not exactly uniform – indeed, when $j = k$, it is easily distinguishable from uniform unless N happens to be very close to a power of 2.

[5] Even under MVW, the result of [19] requires that $\rho \ge \log N - \frac{1}{2}\log(e) + \lambda + O(1)$. Combined with the requirement that $\log(e) \le \frac{1}{2}\log(p) - \lambda$, we get a message of length $m = \log(N) - \rho - O(1) \le \frac{1}{8}\log(N) - \frac{3}{2}\lambda - O(1)$.

Depending on the application, it may be important to have $x[i:j]$ indistinguishable from a truly uniform string. In that case, one may either set $i = 1$ (use the least significant bits) or, in the case $j = k$, ignore the top $\log(1/\epsilon)$ bits of $r[i;k]$ (effectively reducing the number of hardcore bits to about $\log(e) - 3\log(1/\epsilon)$ bits).

Proof of Theorem 9. We define two games. Let $U \leftarrow\!\!\$\ \mathbb{Z}_N^*$. $\mathbf{Game}_0$ is to distinguish $(N, e, x^e \bmod N, x[i,j])$ and $(N, e, U^e \bmod N, x[i,j])$; $\mathbf{Game}_1$ is to distinguish $(N, e, U^e \bmod N, x[i,j])$ and $(N, e, U^e \bmod N, r)$. Since x is chosen uniform randomly from $\mathbb{Z}_N^*$, the advantage in $\mathbf{Game}_1$ is at most 2^{j-k} (since k is the bit length). Let $\mathcal{D}$ be any distinguisher, and let $\overline{\mathcal{D}}$ be the distinguisher for the Φ-Hiding game that prepares inputs to $\mathcal{D}$ using a challenge public key and uses $\mathcal{D}$'s output as its own. We have

$$\begin{aligned} Adv_{\mathcal{D}}^{\mathbf{Game}_0}(1^\lambda) &= |\Pr[\mathcal{D}(N, e, x^e \bmod N, x[i,j]) = 1] - \Pr[\mathcal{D}(N, e, (\mathbb{Z}_N^*)^e \bmod N, x[i,j]) = 1]| \\ &\leq Adv_{c,\theta,\overline{\mathcal{D}}}^{\Phi A}(\lambda) + SD(\mathcal{P}^e \bmod N, U^e \bmod N) \end{aligned}$$

where $\mathcal{P}$ is the set of integers with bits i through j set to $x[i:j]$.

The structure of $\mathcal{P}$ depends on the integers i and j. In general, when $j < k$ and $i > 1$, $\mathcal{P}$ may not be well-approximated by an arithmetic progression. However, if $j = k$, then $\mathcal{P}$ is the arithmetic progression $\mathcal{P} = \{x[i,j] \cdot 2^{i-1} + a \mid a = 0, \ldots, 2^{i-1} - 1\}$. If $i = 1$, then the set $\mathcal{P}$ is more complicated, but it is closely approximated by an AP. Specifically, let $\mathcal{P}' = \{x[i,j] + b \cdot 2^j \mid b = 0, ..., N_j\}$, where $N_j \stackrel{def}{=} N \text{ div } 2^j$ is the integer obtained by consider only bits $j+1$ through k of the binary representation of the modulus N. Then the uniform distribution on $\mathcal{P}$ is at most 2^{k-j}-far from the uniform distribution on $\mathcal{P}'$.

As Theorem 2 applies to arithmetic progressions, we can apply it in the cases $i = 1$ and $j = k$. By Theorem 2,

$$Adv_{\mathcal{D}}^{\mathbf{Game}_0}(1^\lambda) \leq 2\left(\tfrac{2p}{q-1} + \sqrt{\tfrac{N}{e2^{k-|j-i|}}}\right) < 2\epsilon\,.$$

The last inequality uses the hypotheses that $\theta = \log q - \log p \geq 4 + \log\frac{1}{\epsilon}$ and $|j - i| < \log e - 2\log\frac{1}{\epsilon} - 2$. □

Concrete Parameters: Let λ denote the security parameter. As in the calculations for PKCS in the previous section, we require $\log(e) \leq \frac{\log p}{2} - \lambda$ (for Coppersmith's attack to be ineffective) and $\epsilon = 2^{-\lambda}$. To apply Theorem 9, we require that $\theta \geq 4 + \log\frac{1}{\epsilon} = 4 + \lambda$, and therefore $\log e \leq \frac{k-\theta}{4} - \lambda \leq \frac{k-5\lambda}{4} - 1$. Theorem 9 then proves security for a run of bits with length $\log e - 2\lambda - 2 = \frac{1}{4}k - \frac{13}{4}\lambda - 3$. For example, for a modulus of length $k = 2048$ bits and security parameter $\lambda = 80$, we get that the 249 least significant bits are simultaneously hardcore. Alternatively, our analysis shows that the 169 bits in positions $k - 249$ through $k - 169$ are simultaneously hardcore (see the discussion immediately after Theorem 9).

References

[1] A. Akavia, S. Goldwasser, and S. Safra. Proving hard-core predicates using list decoding. In *44th Annual Symposium on Foundations of Computer Science*, pages 146–159. IEEE Computer Society Press, Oct. 2003.

[2] W. Alexi, B. Chor, O. Goldreich, and C. Schnorr. RSA and Rabin functions: Certain parts are as hard as the whole. *SIAM Journal on Computing*, 17(2):194–209, Apr. 1988.

[3] M. Bellare and P. Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In V. Ashby, editor, *ACM CCS 93: 1st Conference on Computer and Communications Security*, pages 62–73. ACM Press, Nov. 1993.

[4] D. Bleichenbacher. Chosen ciphertext attacks against protocols based on the RSA encryption standard PKCS #1. In H. Krawczyk, editor, *Advances in Cryptology – CRYPTO'98*, volume 1462 of *Lecture Notes in Computer Science*, pages 1–12. Springer, Aug. 1998.

[5] M. Blum and S. Goldwasser. An efficient probabilistic public-key encryption scheme which hides all partial information. In G. R. Blakley and D. Chaum, editors, *Advances in Cryptology – CRYPTO'84*, volume 196 of *Lecture Notes in Computer Science*, pages 289–302. Springer, Aug. 1985.

[6] C. Cachin. Efficient private bidding and auctions with an oblivious third party. In *ACM CCS 99: 6th Conference on Computer and Communications Security*, pages 120–127. ACM Press, Nov. 1999.

[7] C. Cachin, S. Micali, and M. Stadler. Computationally private information retrieval with polylogarithmic communication. In J. Stern, editor, *Advances in Cryptology – EUROCRYPT'99*, volume 1592 of *Lecture Notes in Computer Science*, pages 402–414. Springer, May 1999.

[8] D. Coppersmith. Small solutions to polynomial equations, and low exponent RSA vulnerabilities. *Journal of Cryptology*, 10(4):233–260, 1997.

[9] J.-S. Coron, M. Joye, D. Naccache, and P. Paillier. New attacks on PKCS#1 v1.5 encryption. In B. Preneel, editor, *Advances in Cryptology – EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 369–381. Springer, May 2000.

[10] C. Gentry, P. D. Mackenzie, and Z. Ramzan. Password authenticated key exchange using hidden smooth subgroups. In V. Atluri, C. Meadows, and A. Juels, editors, *ACM CCS 05: 12th Conference on Computer and Communications Security*, pages 299–309. ACM Press, Nov. 2005.

[11] O. Goldreich. *Foundations of Cryptography: Basic Applications*, volume 2. Cambridge University Press, Cambridge, UK, 2004. ISBN ISBN 0-521-83084-2 (hardback).

[12] D. Heath-Brown and S. Konyagin. New bounds for gauss sums derived from kth powers, and for heilbronn's exponential sum. *The Quarterly Journal of Mathematics*, 51(2):221–235, 2000.

[13] B. Hemenway and R. Ostrovsky. Public-key locally-decodable codes. In D. Wagner, editor, *Advances in Cryptology – CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 126–143. Springer, Aug. 2008.

[14] D. Hofheinz and E. Kiltz. Practical chosen ciphertext secure encryption from factoring. In A. Joux, editor, *Advances in Cryptology – EUROCRYPT 2009*, volume 5479 of *Lecture Notes in Computer Science*, pages 313–332. Springer, Apr. 2009.

[15] S. Hohenberger and B. Waters. Short and stateless signatures from the RSA assumption. In S. Halevi, editor, *Advances in Cryptology – CRYPTO 2009*, volume 5677 of *Lecture Notes in Computer Science*, pages 654–670. Springer, Aug. 2009.

[16] S. A. Kakvi and E. Kiltz. Optimal security proofs for full domain hash, revisited. In *EUROCRYPT*, pages 537–553, 2012.

[17] J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. Chapman and Hall/CRC Press, 2007.

[18] E. Kiltz, A. O'Neill, and A. Smith. Instantiability of RSA-OAEP under chosen-plaintext attack. In *CRYPTO*, 2010.

[19] M. Lewko, A. O'Neill, and A. Smith. Regularity of lossy RSA on subdomains and its applications. In *EUROCRYPT*, 2013.

[20] A. May. Using lll-reduction for solving rsa and factorization problems. In B. V. Phong Q. Nguyen, editor, *The LLL Algorithm: Survey and Applications*, pages 315–348. Springer, 2010.

[21] H. Montgomery, R. Vaughan, and T. Wooley. Some remarks on gauss sums associated with kth powers. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 118, pages 21–33. Cambridge Univ Press, 1995.

[22] C. Peikert and B. Waters. Lossy trapdoor functions and their applications. In C. Dwork, editor, *STOC*, pages 187–196. ACM, 2008. ISBN 978-1-60558-047-0.

[23] R. L. Rivest, A. Shamir, and L. M. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126, 1978.

[24] C. Schridde and B. Freisleben. On the validity of the phi-hiding assumption in cryptographic protocols. In J. Pieprzyk, editor, *Advances in Cryptology – ASIACRYPT 2008*, volume 5350 of *Lecture Notes in Computer Science*, pages 344–354. Springer, Dec. 2008.

[25] R. Steinfeld, J. Pieprzyk, and H. Wang. On the provable security of an efficient RSA-based pseudorandom generator. In X. Lai and K. Chen, editors, *Advances in Cryptology – ASIACRYPT 2006*, volume 4284 of *Lecture Notes in Computer Science*, pages 194–209. Springer, Dec. 2006.

A Lemma 6

Proof of Lemma 6. This proof is observed by [19]. However, in [19], $\omega - 1$ may not be invertible in $\mathbb{Z}_N$ (recall that $\omega \in \{x | x^e \bmod N = 1\}$) but $\omega - 1$ is invertible in $\mathbb{Z}_p$ as $e | p - 1$.

Let $\mathcal{Q}$ be the distribution of $(V, (V + X)^e \bmod p)$ and $\mathcal{T}$ be the distribution of $(V, U^e \bmod p)$. $\mathcal{Q}_0$ is identical to $\mathcal{Q}$ except that the event $(V + X)^e \bmod p = 0$ occurs; $\mathcal{T}_0$ is identical to $\mathcal{T}$ except that the event $U^e \bmod p = 0$ occurs. Similarly, $\mathcal{Q}_1$ is defined to be identical to $\mathcal{Q}$ except that $(V + X)^e \bmod p \neq 0$; $\mathcal{T}_1$ is identical to $\mathcal{T}$ except that $U^e \bmod p \neq 0$. Then, we have:

$$SD(\mathcal{Q}, \mathcal{T}) = SD(\mathcal{Q}_0, \mathcal{T}_0) + SD(\mathcal{Q}_1, \mathcal{T}_1).$$

$$
\begin{aligned}
SD(\mathcal{Q}_0, \mathcal{T}_0) \leq & < 1, \mathcal{Q}_0 > + < 1, \mathcal{T}_0 > \\
& \leq \frac{1}{p-1} + \frac{1}{p-1} \leq \frac{2}{p-1}.
\end{aligned}
$$

$$
\begin{aligned}
SD(\mathcal{Q}_1, \mathcal{T}_1) &\leq \sqrt{supp(\mathcal{Q}_1 - \mathcal{T}_1)||\mathcal{Q}_1||_2^2 - 1} \\
&\leq \sqrt{\frac{(p-1)^2}{e}||\mathcal{Q}_1||_2^2 - 1}.
\end{aligned}
$$

Where,

$$
\begin{aligned}
||\mathcal{Q}_1||_2^2 &= \Pr[(V, (V+X)^e \bmod p) = (V', (V'+Y) \bmod p)] \\
&= \frac{1}{p-1}\Pr[(V+X)^e \bmod p = (V+Y)^e \bmod p] \\
&= \frac{1}{p-1} \sum_{\omega \in \{x | x^e \bmod p = 1\}} \Pr[(V+X) = \omega(V+Y) \bmod p] \\
&= \frac{1}{p-1}(\Pr[X = Y \bmod p] + \sum_{\omega \neq 1} \Pr[V = (\omega-1)^{-1}(X - \omega Y) \bmod p]) \\
&= \frac{1}{p-1}(\Pr[X = Y \bmod p] + \frac{e-1}{p-1}) \\
&\leq \frac{1}{p-1}(\frac{1}{p} + \frac{1}{\overline{K}} + \frac{e-1}{p-1}) \\
&\leq \frac{1}{p-1}(\frac{e}{p} + \frac{1}{\overline{K}}).
\end{aligned}
$$

Therefore, we have:

$$
\begin{aligned}
& SD(\mathcal{Q}, \mathcal{T}) \\
& \leq \frac{2}{p-1} + \sqrt{\frac{p-1}{e}(e/p) - 1 + \frac{p-1}{e\overline{K}}} \\
& \leq \frac{2}{p-1} + \sqrt{\frac{p-1}{e\overline{K}}}.
\end{aligned}
$$

□