Public-Key Encryption with Weak Randomness: Security against Strong Chosen Distribution Attacks

Damien Vergnaud[*] David Xiao[†]

October 23, 2013

Abstract

Chosen Distribution Attacks (CDA) were introduced by Bellare *et al.* (Asiacrypt '09) to model attacks where an adversary can control the distribution of both messages and random coins used in an encryption scheme. One important restriction in their definition is that the distributions chosen by the adversary cannot depend on the public key being attacked, and they show that some restriction of this form is necessary (for the same reasons that secure deterministic encryption is impossible if we allow arbitrary dependence between the plaintext distributions and the public key).

Subsequently Raghunathan *et al.* (Eurocrypt '13) showed how to relax this restriction by allowing the message/randomness distributions to depend on the public key as long as the distributions belong to a family of bounded size fixed before the public key is known.

We extend the definition further to what we call Strong Chosen Distribution Attacks where the message/randomness distributions may depend on the public key as long as certain entropy conditions are satisfied. Our security model comes from a natural model of attack where an adversary infiltrates the encryption system and installs a trojan program prior to knowing the public key, and subsequently is allowed limited communication with the trojan program.

We present secure constructions in the standard and random oracle models both with and without decryption oracles (corresponding to CPA or CCA security). We also prove that our definition simultaneously generalizes previous definitions in this line of work.

[*] Ecole Normale Supérieure. E-mail: damien.vergnaud@ens.fr

[†] CNRS, LIAFA, Université Paris 7. E-mail: dxiao@liafa.univ-paris-diderot.fr

1 Introduction

The classical notion of semantic security for public-key encryption schemes [GM84] asks that it must be infeasible for a computationally-bounded adversary to compute any information about a plaintext from its ciphertext and the corresponding public encryption key. This notion can only be reached by a probabilistic encryption scheme.

In theoretical models, the process of generating randomness is generally assumed to behave perfectly, giving the users access to sequences of independent and uniform bits. Unfortunately, it turns out to be extremely difficult to obtain perfect randomness in practice. Indeed, the physical sources that produce randomness often have non-uniform or even unknown distributions. Moreover, even if some processing is done on the randomness in order to give it a better quality, there will be many scenarios in which an attacker could be able to recover some useful information.

For several probabilistic encryption schemes, using weak randomness can lead to catastrophic attacks in particular when the used randomness has little or even no min-entropy. For instance, ElGamal encryption scheme [Gam85] and all hybrid encryption schemes are vulnerable to plaintext recovery attacks when the randomness used is predictable. It seems therefore desirable to construct encryption schemes in such a way that using weak randomness will have as little as possible impact on the security of the scheme. In [BBN$^+$09], Bellare, Brakerski, Naor, Ristenpart, Segev, Shacham and Yilek introduced the idea to provide two tiers of security: with good randomness the scheme achieves the classical semantic security notion but when it is fed with bad randomness, it achieves some weaker but still useful notion of security.

Formally, Bellare *et al.* asked that security is guaranteed as long as the joint distribution of the message and randomness (chosen by the adversary) has sufficiently high min-entropy and called this notion *Indistinguishability under a Chosen Distribution Attack* (CDA). Their definition is inspired by the one for deterministic encryption from [BBO07] and one important restriction in their definition is that the message/randomness distribution cannot depend on the public key being attacked. For this reason, we will call their notion *weak* CDA. In particular, weak CDA does not encompass standard semantic security. One way to view their model is that there is a "sharp" threshold between when randomness is perfect (where messages can depend arbitrarily on the public key) and when randomness is imperfect (where there can be *no* dependence between the message/randomness distribution and the public key).

Bellare *et al.* proposed several constructions from any public-key encryption in the random oracle model. Following previous work on deterministic encryption [BFOR08, BFO08], they also proposed a construction in the standard model relying on the notion of lossy trapdoor function introduced by Peikert and Waters [PW11]. Subsequent work on deterministic encryption also implies analogous strenghened results on weak CDA security [BS11, FOR12, Wee12], where those papers consider auxiliary input and correlated message distributions.

Recently, Raghunathan, Segev and Vadhan [RSV13b] showed how to relax the restriction that the message/randomness distribution be independent of the public key. They allow the adversary to pick a family of distributions before seeing the public key, where the family must be of bounded size, and then they permit the message/randomness distribution used by the adversary to be chosen depending on the public key, as long as the distribution belongs to this previously fixed family. In a related work, Birrel *et al.* [BCPT13] study the question of randomness-dependent message security where the adversary can select the message to be encrypted depending both on the random coins and the public key; they show positive results when one can assume that the amount of dependence on the random coins is limited.

The goal of this paper is to extend the definition further to what we call *Strong Chosen Distribution Attacks* where the message/randomness distributions may depend on the public key as long as certain entropy conditions are satisfied. In particular, we will present a definition that simultaneously generalizes the previous definitions, *i.e.* a scheme secure in our model will be simultaneously secure against chosen plaintext attacks, (weak) chosen distribution attacks, randomness-dependent messages, etc. In particular, we provide a "smooth" tradeoff between perfect and imperfect randomness: when the randomness is perfect the message can depend arbitrarily on the public key, while security is preserved even when the message/randomness have imperfect entropy, and the amount of dependence on the public key degrades in a smooth, graceful manner rather than in a sharp way as in [BBN$^+$09].

1.1 Our model and definition

Our adversary is divided in several components and one may conceptually view its attack as proceeding in several phases in sequence:

Infiltration The adversary installs an interactive algorithm Rand on the target system. This occurs prior to any information about the public key being revealed. The program Rand is a trojan that sits between the target system's encryption function and its source of randomness: each time the target system tries to encrypt a message, the trojan feeds it some (potentially adversarially biased) random coins. Furthermore, Rand is *interactive*, and so in later phases it may communicate with an outside system who may exert control over the distribution of the modified random coins.

Revelation The adversary is allowed to learn some (possibly limited) amount of information $\mathsf{Rev}(\mathrm{pk})$ about the public key.

Query The adversary creates an interactive message generating algorithm Mesg that, using (possibly limited) communication with the Rand algorithm installed in the infiltration phase, generates a message distribution (that may be correlated to the randomness distribution generated by Rand because of the communication). Samples drawn from this message/randomness distribution are then encrypted by the target system and sent to the adversary. The adversary may adaptively use the results of these encryptions to generate a new Mesg and repeat the attack to see more ciphertexts, and this may be repeated a polynomial number of times.

Distinguishing The adversary runs a distinguisher Dist and guesses whether the ciphertexts he saw during the query phase were encryptions of message-randomness pairs drawn from the distribution he created, or whether they are encryptions of independent and uniform message-randomness pairs. In this phase the distinguisher may know the entire public key.

In the above scenario, it is known that if the message-randomness distribution may depend arbitrarily on the public key, then security is impossible [BBN$^+$09, BBO07]. Our main contribution is a set of conditions that are both necessary and sufficient, and also simultaneously generalizes all of the previous definitions studied in the literature [GM84, BBN$^+$09, BCPT13, RSV13a]. Informally speaking, we use the following three entropy conditions:

1. From the point of view of the randomness-generating algorithm Rand, there must be significant entropy in the public key.
2. From the point of view of message-generating algorithm Mesg, there must be significant entropy in the ciphertext (this may come either from the public key, or from the random coins used in the encryption).

3. Knowing the code of the algorithms $\mathsf{Mesg}, \mathsf{Rand}$, there must still be significant entropy in the messages they produce.

At an intuitive level it is clear that these three conditions are necessary: the first condition is necessary because if Rand knows the public key then it can always generate random coins so that a plaintext of a single bit b encrypts to, say, a ciphertext whose first bit is b (this is possible assuming the ciphertexts look uniform; if not, one can still show similar attacks by using hash functions to randomize the predicate that will reveal b). In a similar fashion, the second condition is necessary because if Mesg knows both the public key and the random coins to be used, it can choose its message m so that the first bit of the m equals, say, the first bit of the encryption of m. The third condition is necessary because otherwise an attacker would know with significant confidence the message/randomness used to generate a ciphertext, and (knowing the public key) can just compute the encryption itself to see whether the resulting ciphertext matches the one that it is supposed to distinguish.

We show that using an appropriate formalization of these three conditions (see Definition 3.2), they turn out to be *sufficient* as well. At a high level, to achieve the first entropy condition we restrict the communication that may flow from the public-key revelation algorithm to the randomness-generating algorithm Rand. This is well-motivated by reality, since communication from the trojan program to the outside world may be noticed by the target system if there is too much of it, while a small amount of communication may pass unnoticed. To achieve the second entropy condition we require that, conditioned on the transcript of the interaction between $\mathsf{Mesg}, \mathsf{Rand}$, there remains large entropy in the randomness that Rand outputs. To achieve the third entropy condition, we require that conditioned on what an observer can see, there remains large entropy in the message/randomness generated by the interaction of $\mathsf{Mesg}, \mathsf{Rand}$.

1.2 Our results

We construct schemes in the standard and random oracle models that satisfy our new definition of Strong Chosen Distribution Attack security. As described above, in our notion of security we require that certain random variables including the public key, the message distribution, and the random coin distribution each have significant entropy from the attacker's point of view. Our random oracle construction has essentially optimal parameters, *i.e.* it only requires that public key, randomness, and messages have super-logarithmic entropy from the adversary's point of view, which is the best one can hope for. Our standard model construction has good parameters except for the public key, and improving this dependence is an interesting open question; we discuss the quality of the parameters of our standard model construction in Section 6.

Our random-oracle model construction follows the well-known "encrypt-with-hash" paradigm [BBO07], where we take an existing encryption scheme and modify it so that instead of feeding it random coins directly, we use the random oracle applied to the public key, message, and random coins to obtain "hashed" random coins that we then use in the original encryption. We note that to achieve our security notion we will require that the original scheme be not only semantically secure but also *anonymous* [BBDP01], *i.e.* the ciphertexts do not leak information about the public key being used; this is similar to the anonymity required in the adaptively secure constructions of [BBN+09].

Our standard-model constructions are based on the scheme of Raghunathan *et al.* [RSV13a], namely we apply a t-wise δ-dependent permutation to the message and random coins, and then apply a lossy trapdoor function. However, in order to prove security we cannot use their analysis directly. Roughly, the main technical lemma of [RSV13a] shows that applying a t-wise δ-dependent permutation and then a lossy function is a good extractor with high probability for *all sources simultaneously* in some family of high-entropy sources, where the family is of bounded size. In our application, it turns out that the family of sources is much too large to be handled by the theorems of [RSV13a] and so we prove stronger versions of their technical lemmas relying on the particular structure of the sources we encounter, which will in turn prove security for our scheme. In addition, we improve on the key length in the CCA scheme as compared to [RSV13a] (see Section 5.3 for details).

We also prove that our definition simultaneously generalizes all the previous definitions in this line of work, including the definition of Raghunathan *et al.* [RSV13a], the (weak) chosen-distribution-attack model and hedged encryption model of Bellare *et al.* [BBN^{+}09], and the randomness-dependent message security model of Birrell *et al.* [BCPT13]. (These models in turn generalize previous notions such as those of [MS09, HO10].) Thus, our definition is the strongest achievable definition to date in this line of work.

2 Preliminaries

For a positive integer i we define $[i] = \{1, \dots, i\}$. Let S be any set and T be any positive integer; for a vector $v \in S^T$ we let $v_{[i]}$ denote the sub-vector in S^i given by $v_{[i]} = (v_1, \dots, v_i)$. For vectors $m \in (\{0,1\}^n)^T, r \in (\{0,1\}^\ell)^T$, we let $(mr)_i = (m_i, r_i)$ and $(mr)_{[i]} = ((m_1, r_1), \dots, (m_i, r_i))$. For strings $m_i, r_i \in \{0,1\}^*$ we let $m_i \parallel r_i$ or simply $m_i r_i$ denote their concatenation.

We let bold-faced variables such as $\mathbf{x}$ denote random variables, while normal variables such as x denote a particular value that $\mathbf{x}$ may take. We let $\mathrm{supp}(\mathbf{x})$ denote its support, *i.e.* the set of elements on which $\mathbf{x}$ has non-zero probability. All logarithms are base 2 unless otherwise stated. The min-entropy function is defined as $\mathrm{H}_\infty(\mathbf{x}) = \min_{x \in \mathrm{supp}(\mathbf{x})} \log(1/\Pr[\mathbf{x} = x])$. Let $\mathbf{m}$ be a random variable over S^T, then we say that $\mathbf{m}$ is a κ source if for all $i \in [T]$, $\mathrm{H}_\infty(\mathbf{m}_i) \geq \kappa$, and we say that $\mathbf{m}$ is a κ-block-wise-source if for all $i \in [T], m_{[i-1]} \in \mathrm{supp}(\mathbf{m}_{[i-1]})$ it holds that $\mathrm{H}_\infty(\mathbf{m}_i \mid \mathbf{m}_{[i-1]} = m_{[i-1]}) \geq \kappa$. The statistical distance between two random variables $\mathbf{X}$ and $\mathbf{Y}$ over a finite domain Ω is $\Delta(\mathbf{X}, \mathbf{Y}) = \frac{1}{2} \sum_{\omega \in \Omega} |\Pr[\mathbf{X} = \omega] - \Pr[\mathbf{Y} = \omega]|$. A function $\varepsilon : \mathbb{N} \to [0,1]$ is negligible if $\varepsilon(k) = k^{-\omega(1)}$ and is overwhelming if $\varepsilon(k) = 1 - k^{-\omega(1)}$.

We will use the following standard lemma [Vad12, Lemma 6.30]:

Lemma 2.1 (Chain rule for min-entropy)**.** *Let* $\mathbf{x}, \mathbf{y}$ *be random jointly distributed random variables, and such that* $|\mathrm{supp}(\mathbf{x})| = s$*. Then for all* $t \geq 0$*, it holds that* $\Pr_{y \leftarrow_R \mathbf{y}}[\mathrm{H}_\infty(\mathbf{x} \mid \mathbf{y} = y) < \mathrm{H}_\infty(\mathbf{x}) - \lfloor \log s \rfloor - t] \leq 2^{-t}$.

2.1 Lossy trapdoor functions

Lossy trapdoor functions were first defined in [PW11] and have found numerous applications in the recent years. They consists of two families of functions: functions in one family are injective and can be efficiently inverted using a trapdoor while functions in the other family are ”lossy” (*i.e.* the size of their image is significantly smaller than the size of their domain). One requires that the description of randomly chosen function from the family of injective functions and from the family of lossy functions are computationally indistinguishable.

Definition 2.2. A family of (n, a)-lossy trapdoor functions is defined by two probabilistic polynomial-time algorithms $\mathsf{LGen}_0, \mathsf{LGen}_1$ satisfying the following properties:

1. $\mathsf{LGen}_0(1^k)$ samples a circuit f_0 taking $n = n(k)$ bits of input and producing $n' \geq n$ bits of output.

We identify f_0 with the function $\{0,1\}^n \to \{0,1\}^{n'}$ that it computes. It holds that the image of f_0 has size at most 2^a.

2. $\mathsf{LGen}_1(1^k)$ samples a pair of circuits (f_1, g_1) where f_1 takes n bits of input and produces n' bits of output, while g_1 takes n' bits of input and produces n bits of output, and for all $x \in \{0,1\}^n$, it holds that $g_1(f_1(x)) = x$.

3. The two random variables $\mathbf{f}_0$ and $\mathbf{f}_1$ are computationally indistinguishable.

The notion of $\mathcal{R}$-lossy trapdoor function family was introduced by Raghunathan, Segev and Vadhan in [RSV13a]. It generalizes the previous notion in such a way that instead of having only two branches (a lossy branch and an injective branch) they have many branches, some of which are injective and some of which are lossy. To achieve this property, the $\mathcal{R}$-lossy trapdoor functions take as input an additional argument, called a *tag*, which is a binary string of appropriate length with no particular structure. The tags are partitioned into two subsets: injective tags, and lossy tags.

The partitioning of the tags is defined by a binary relation $\mathcal{R} \subset \mathcal{K} \times \mathcal{T}$ and the key-generation algorithm receives as input an initialization $K \in \mathcal{K}$ that partitions the set of tags $\mathcal{T}$ so that $t \in \mathcal{T}$ is lossy if and only if $(K,t) \in \mathcal{R}$.

Definition 2.3. A family of $\mathcal{R}$-(n,a)-lossy trapdoor functions is defined by a probabilistic polynomial-time algorithms KGen which takes as input a key $K \in \mathcal{K}$ and samples a pair of circuits (f,g) where f takes n bits of input and a tag $t \in \mathcal{T}$ and produces n' bits of output, while g takes n' bits of input and a tag $t \in \mathcal{T}$ and produces n bits of output, such that:

1. if $(K,t) \in \mathcal{R}$, $f(\cdot,t)$ has size at most 2^a.

2. if $(K,t) \notin \mathcal{R}$, it holds that $g(f(x,t),t) = x$ for all $x \in \{0,1\}^n$.

3. For every polynomial time adversary $\mathcal{A}$, there is a negligible function $\mathsf{negl}(k)$ such that

$$\mathrm{Adv}^{\mathsf{KGen}}(\mathcal{A},k) := |\Pr[\mathcal{R}\text{-}\mathsf{Lossy}_0^{\mathsf{KGen}}(\mathcal{A},k) = 1] - \Pr[\mathcal{R}\text{-}\mathsf{Lossy}_1^{\mathsf{KGen}}(\mathcal{A},k) = 1]| \leq \mathsf{negl}(k).$$

where for $b \in \{0,1\}$, the random experiment $\mathcal{R}\text{-}\mathsf{Lossy}_b^{\mathsf{KGen}}(\mathcal{A},k)$ is define as follows:

$$\begin{array}{lll} \mathcal{R}\text{-}\mathsf{Lossy}_b^{\mathsf{KGen}}(\mathcal{A},k) & := & (K_0, K_1, \mathsf{state}) \xleftarrow{R} \mathcal{A}_1(1^k) \\ & & (f,g) \xleftarrow{R} \mathsf{KGen}(1^k, K_b) \\ & & b' \xleftarrow{R} \mathcal{A}_2(f, \mathsf{state}) \\ & & \text{Output } b' \end{array}$$

These functions turn out to be particularly suitable to constructing encryption schemes secure against chosen-ciphertext attacks where an adversary is given access to a decryption oracle. In this scenario, the security reduction must be able to answer the decryption queries of the adversary but should not be able to obtain any unknown information from the challenge ciphertext. An approach to reach this security level is to (secretly) partition the ciphertext space into two subsets so that in the security proof the decryption queries corresponds to injective tags while the challenge ciphertext corresponds to a lossy tag (with non-negligible probability).

The notion of admissible hash functions, introduced to prove the full security of some identity-based encryption schemes, was introduced by Boneh and Boyen in [BB04] to realize this secret partition.

Let $\nu \in \mathbb{N}$. For $K \in \{0,1,\perp\}^\nu$, we define the partitioning function $P_K : \{0,1\}^\nu \to \{\mathtt{Lossy}, \mathtt{Inj}\}$ as follows:

$$P_K(y) := \begin{cases} \mathtt{Lossy} & \text{if } K_i \in \{y_i, \perp\}, \forall i \in \{1, \ldots, \nu\} \\ \mathtt{Inj} & \text{otherwise.} \end{cases}$$

For any $u < \nu$, we denote $\mathcal{K}_{u,\nu}$ the uniform distribution over $\{0,1,\perp\}^\nu$ conditioned on exactly u position having $\perp$ values (such that for any K sampled according to $\mathcal{K}_{u,\nu}$, $\#\{y \in \{0,1\}^\nu, P_K(y) = \mathtt{Lossy}\} = 2^u$).

Definition 2.4 (Admissible hash functions)**.** Let $\mathcal{H} = \{\mathcal{H}_k\}_{k\in\mathbb{N}}$ be a hash-function ensemble where each $h \in \mathcal{H}_k$ is a polynomial-time computable $h : \{0,1\}^{n(k)} \to \{0,1\}^{\nu(k)}$. We say that $\mathcal{H}$ is an admissible hash-function ensemble if for every $h \in \mathcal{H}$, there exists an efficiently recognizable set $\mathsf{Unlikely}_h \subset \cup_{q\in\mathbb{N}}(\{0,1\}^{n(k)})^q$ of string tuples such that the following two properties hold:

1. For every PPT algorithm $\mathcal{A}$, there exists a negligible function $\mathsf{negl}(k)$ satisfying

$$\Pr[(x_0, \ldots, x_q) \in \mathsf{Unlikely}_h] \leq \mathsf{negl}(k)$$

   where $h \xleftarrow{R} \mathcal{H}_k$ and $(x_0, \ldots, x_q) \xleftarrow{R} \mathcal{A}(1^k, h)$.

2. For every polynomial $q = q(k)$, there exists a polynomial $\Gamma = \Gamma(k)$ and an efficiently computable $u = u(k)$ such that, for every $h \in \mathcal{H}_k$ and $(x_0, \ldots, x_{q(k)}) \notin \mathsf{Unlikely}_h$ with $x_0 \notin \{x_1, \ldots, x_q\}$, we have

$$\Pr_{K \xleftarrow{R} \mathcal{K}_{u,\nu}}[P_K(h(x_0)) = \mathtt{Lossy} \wedge P_K(h(x_1)) = \cdots = P_K(h(x_{q(k)})) = \mathtt{Inj}] \geq \frac{1}{\Gamma(k)}.$$

Boneh and Boyen [BB04] proved that the existence of admissible hash function is implied by collision-resistant hash functions.

We will use a collection of $\mathcal{R}$-lossy trapdoor function in conjunction with admissible hash function where the relation $\mathcal{R} = \mathcal{R}^{BM}$ is defined by a key $K \in \{0,1,\perp\}^\nu$ such that for $t \in \{0,1\}^\nu \in \mathcal{R}^{BM}$ if and only if $K_i \in \{t_i, \perp\}, \forall i \in \{1, \ldots, \nu\}$ (*i.e.* $P_K(t_i) = \mathtt{Lossy}$).

2.2 t-wise δ-dependent permutations

Definition 2.5. Let $\boldsymbol{\pi}$ be a random variable over permutations $\{0,1\}^n \to \{0,1\}^n$. We say that $\boldsymbol{\pi}$ is a t-wise δ-dependent permutations if for all $x_1, \ldots, x_t \in \{0,1\}^n$, it holds that

$$\Delta((\boldsymbol{\pi}(x_1), \ldots, \boldsymbol{\pi}(x_t)), (\boldsymbol{\pi}^*(x_1), \ldots, \boldsymbol{\pi}^*(x_t))) \leq \delta$$

where $\boldsymbol{\pi}^*$ is a truly random permutation $\{0,1\}^n \to \{0,1\}^n$.

There exist efficiently constructible t-wise δ-dependent permutations with key length $O(nt + \log(1/\delta))$ and that are efficiently computable and invertible given the key [KNR09].

2.3 Crooked leftover hash lemma

We will strengthen the following "average-case" crooked leftover hash lemma proved by [RSV13a]:

Lemma 2.6 ([RSV13a])**.** *Let $f : \{0,1\}^n \to \{0,1\}^{n'}$ be (n,a)-lossy and $\boldsymbol{\pi}$ be a t-wise δ-dependent permutation over $\{0,1\}^n$ satisfying $t \geq 8$ even and $\delta \leq 2^{-nt}$. For all $\varepsilon \in (0,1)$ and random variables $(\mathbf{x},\mathbf{y})$ over $\{0,1\}^m \times \{0,1\}^n$ such that for all $x \in \mathrm{supp}(\mathbf{x})$ it holds that*

$$\mathrm{H}_\infty(\mathbf{y} \mid \mathbf{x} = x) \geq a + 2\log(1/\varepsilon) + 2\log t + \Theta(1)$$

Then with probability $\geq 1 - 2^{a-t+1}/\varepsilon$ over the choice of $\pi \overset{R}{\leftarrow} \boldsymbol{\pi}$ it holds that

$$\Delta((\mathbf{x}, f(\pi(\mathbf{y}))), (\mathbf{x}, f(\mathbf{u}_n))) \leq \varepsilon$$

3 Strong Chosen Distribution Attack Security

An intuitive view of and motivation for our definition were presented in Section 1.1. We now proceed formally to define our model. Define the "real-or-random" oracle RoR as follows. The oracle is parameterized by pk, $\mathsf{mode} \in \{\mathsf{real}, \mathsf{random}\}$, and an integer T given in unary. The RoR oracle takes input two interactive algorithms Mesg and Rand. (In our security game, Rand will be fixed once and for all at the start, while the adversary may query many different Mesg during the course of the security game.)

After executing Mesg and Rand interactively with each other, the output of Rand is a vector of T random bit strings $r_1, \ldots, r_T$ and the output of Mesg is a vector of T messages $m_1, \ldots, m_T$ and an additional "hint" string ξ.

- $\mathsf{RoR}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{Rand}, \mathsf{random}, 1^T)$ runs Mesg and Rand interactively with each other to determine ξ (it throws away all the other outputs). It then samples $(m'_i, r'_i)_{i \in [T]}$ uniformly and independently and outputs $(\xi, \mathsf{Enc}_{\mathrm{pk}}(m'_1; r'_1), \ldots, \mathsf{Enc}_{\mathrm{pk}}(m'_T; r'_T))$.
- $\mathsf{RoR}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{Rand}, \mathsf{real}, 1^T)$ executes Mesg and Rand interactively with each other. Let $\xi, m_1, \ldots, m_T$ be the output of Mesg and $r_1, \ldots, r_T$ be the output of Rand. Output $(\xi, \mathsf{Enc}_{\mathrm{pk}}(m_1; r_1), \ldots, \mathsf{Enc}_{\mathrm{pk}}(m_T; r_T))$.

Definition 3.1 (ROR-SCDA game)**.** Let (Gen, Enc, Dec) be a public-key encryption system with security parameter k. An adversary in the ROR-SCDA game is defined by a sequence of integers $T_k = \mathrm{poly}(k)$ and tuples of circuits $\mathcal{A}_k = (\mathsf{Rand}_k, \mathsf{Rev}_k, \mathsf{Query}_k, \mathsf{Dist}_k)$ for $k \in \mathbb{N}$. The game is defined as follows for $\mathsf{mode} \in \{\mathsf{real}, \mathsf{random}\}$:

$\mathsf{ROR\text{-}SCDA}_{\mathcal{A}}^{\mathsf{mode}}(k)$ game

$(\mathrm{pk}, \mathrm{sk}) \overset{R}{\leftarrow} \mathsf{Gen}(1^k)$

$\tau = \mathsf{Query}_k^{\mathsf{RoR}_{\mathrm{pk}}(\cdot; \mathsf{Rand}_k, \mathsf{mode}, T_k)}(\mathsf{Rev}_k(\mathrm{pk}))$

Output $\mathsf{Dist}_k(\tau, \mathrm{pk})$

$$\mathrm{Adv}_{\mathsf{ROR\text{-}SCDA}}^{\mathcal{A}}(k) = |\Pr[\mathsf{ROR\text{-}SCDA}_{\mathcal{A}}^{\mathsf{real}}(k) = 1] - \Pr[\mathsf{ROR\text{-}SCDA}_{\mathcal{A}}^{\mathsf{random}}(k) = 1]|$$

Mapping Definition 3.1 onto the description of Section 1.1, first the Rand_k algorithm is fixed without knowing pk (infiltration), and then Query_k is given some information about the public key $\mathsf{Rev}_k(\mathrm{pk})$ (revelation). Query_k can make queries to the RoR oracle and output some state τ (query phase). Finally Dist_k gets τ and the entire public key pk and outputs a bit (distinguishing phase).

Define $S(\mathsf{Rand}) = \bigcup_{\mathsf{Mesg}} \mathrm{supp}(\langle \mathsf{Mesg}, \mathsf{Rand} \rangle)$, namely the set of transcripts that Rand could possibly produce when interacting with an arbitrary Mesg. For $\sigma \in S(\mathsf{Mesg}, \mathsf{Rand})$, define $\mathbf{r}_\sigma$ to be the random variable over vectors of random coins distributed according to what Rand would generate conditioned on transcript σ.

Since security is impossible if we allow arbitrary attackers in the ROR-SCDA game, we will restrict our attention to certain profiles of actors, which will be constrained along the lines described in Section 1.1.

Definition 3.2. An attacker $\mathcal{A} = (\mathsf{Rand}, \mathsf{Rev}, \mathsf{Query}, \mathsf{Dist})$ has a *(block-wise) attack profile* $\Pi = (\alpha, \beta, \kappa)$ for a scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ if the following hold. Let ρ denote the output Rev and let σ denote the transcript of communication between Mesg and Rand. Let p denote the length of ρ and c denote the length of σ (we assume *w.l.o.g.* that these lengths are the same over all possible random coins). It must hold that:

1. $\mathrm{H}_\infty(\mathbf{pk}) - \min(p, c) \geq \alpha$.
2. If $p > \mathrm{H}_\infty(\mathbf{pk}) - \alpha$ then for all $\sigma \in S(\mathsf{Rand})$, it must hold that $\mathbf{r}_\sigma$ is a (block-wise) β-source.
3. For any Mesg queried by Query on input $\rho = \mathsf{Rev}(\mathrm{pk})$, let $(\boldsymbol{\xi}, \mathbf{m}, \mathbf{r})$ be the outputs of Mesg and Rand. Viewing $(\mathbf{m}, \mathbf{r})$ as a vector where $(\mathbf{m}, \mathbf{r})_i = (\mathbf{m}_i, \mathbf{r}_i)$, it holds that $(\mathbf{m}, \mathbf{r})$ is a (block-wise) κ-source conditioned on ρ, ξ for all ξ and furthermore for every m in the support of $\mathbf{m}$, it holds that $m_i \neq m_j$ for all $i \neq j \in [T]$.

We say that $\mathcal{A}$ has profile Π with probability $1-\eta$ if the second condition holds for all Mesg with probability $1 - \eta$ over the choice of σ and the third condition holds for all Mesg with probability $1 - \eta$ over the choice of ξ. ($\mathcal{A}$ has profile Π if it has profile Π with probability 1.)

Definition 3.3 (ROR-SCDA security)**.** A scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is defined to be $(\Pi(k), \eta(k), \varepsilon(k, s))$-ROR-SCDA-(block-wise)-secure if $\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ROR\text{-}SCDA}}(k) \leq \varepsilon(k, s)$ for all adversaries $(\mathcal{A}_k)_{k \in \mathbb{N}}$ with circuit size s and having attack profile $\Pi(k)$ with probability $1 - \eta(k)$. It is simply $\Pi(k)$-ROR-SCDA-(block-wise)-secure if for any negligible $\eta(k)$ it is $(\Pi(k), \eta(k), \varepsilon(k, s))$-ROR-SCDA-(block-wise)-secure where $\varepsilon(k, k^c)$ is negligible in k for any constant c.

Definition 3.4 (Chosen ciphertext security)**.** The ROR-SCDCA game (strong chosen distribution and chosen ciphertext security) is identical to the ROR-SCDA game except the adversary (in any of its sub-algorithms) is also allowed access to a Dec' oracle that answers all decryption queries except for the ciphertexts previously output by RoR. (Π, η, ε)-ROR-SCDCA-(block-wise)-security is defined as (Π, η, ε)-ROR-SCDA-(block-wise)-security except using the ROR-SCDCA security game.

Because of the following proposition, when constructing schemes we aim to prove that they are secure for profiles with α, β, κ as small as possible and η as large as possible.

Proposition 3.5. *If a scheme $\mathcal{E}$ is (Π, η, ε)-ROR-SCDA-(block-wise)-secure for some profile $\Pi = (\alpha, \beta, \kappa)$, then it is also $(\Pi', \eta', \varepsilon)$-secure for any $\Pi' = (\alpha', \beta', \kappa')$ where $\alpha' \geq \alpha, \beta' \geq \beta, \kappa' \geq \kappa$ and $\eta' \leq \eta$.*

4 Relation with other security notions

We now prove that (Π, η, ε)-security implies previous notions of security. To make our results as strong as possible, in the following theorems we set $\Pi = (\alpha, \beta, \kappa)$ where α, β, κ are as large as possible and η is as small as possible, since Proposition 3.5 implies that security for smaller values of α, β, κ and larger values of η is even stronger.

4.1 Indistinguishability under CPA attacks

In [GM84], Goldwasser and Micali formalized the notion of *indistinguishability* for public-key encryption schemes which asserts that it must be infeasible for a probabilistic polynomial-time adversary to compute any information about a plaintext from its ciphertext and the corresponding public encryption key. This security notion can be formalized with the following *real-or-random* definition:

Definition 4.1. For every $s : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \times \mathbb{N} \to [0, 1]$, a public-key encryption scheme $\mathcal{E} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is $\varepsilon(\cdot, \cdot)$-IND-CPA secure if for every adversary $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ a pair of circuits of size at most $s(k)$, the ensembles $\{\mathsf{IND\text{-}CPA}^{\mathcal{E}}_{\mathsf{real}}(A, k)\}_{k \in \mathbb{N}}$ and $\{\mathsf{IND\text{-}CPA}^{\mathcal{E}}_{\mathsf{random}}(A, k)\}_{k \in \mathbb{N}}$ are $\varepsilon(k, s(k))$-close where

$$\begin{aligned} \mathsf{IND\text{-}CPA}^{\mathcal{E}}_{\mathsf{mode}}(A, k) \quad := \quad & (\mathrm{pk}, \mathrm{sk}) \overset{R}{\leftarrow} \mathsf{Gen}(1^k) \\ & r \overset{R}{\leftarrow} \mathcal{R} \\ & (m, \mathsf{state}) \overset{R}{\leftarrow} \mathcal{A}_1(1^k, \mathrm{pk}) \\ & \mathsf{if\ mode} = \mathsf{real}, C \leftarrow \mathsf{Enc}_{\mathrm{pk}}(m, r) \\ & \mathsf{if\ mode} = \mathsf{random}, m' \overset{R}{\leftarrow} \mathcal{M}, C \leftarrow \mathsf{Enc}_{\mathrm{pk}}(m', r) \\ & b \overset{R}{\leftarrow} \mathcal{A}_2(C, \mathsf{state}) \\ & \mathsf{Output}\ b \end{aligned}$$

Lemma 4.2. *Let $s : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \times \mathbb{N} \to [0, 1]$. Let $\mathcal{E} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption system with security parameter k, message space $\mathcal{M} = \{0, 1\}^{n(k)}$ and randomness space $\mathcal{R} = \{0, 1\}^{\ell(k)}$. If $\mathcal{E}$ is $(\Pi(\cdot), 0, \varepsilon(\cdot, \cdot))$-ROR-SCDA-(block-wise)-secure with $\Pi(k) = (\mathrm{H}_\infty(\mathbf{pk}), \ell(k), \ell(k))$ then $\mathcal{E}$ is $\varepsilon(\cdot, \cdot)$-IND-CPA-secure.*

Proof. Assume for contradiction that there exists an adversary $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$ that ε-breaks the IND-CPA-security of $\mathcal{E}$.

We will construct an adversary $\mathcal{A} = (\mathsf{Rand}, \mathsf{Rev}, \mathsf{Query}, \mathsf{Dist})$ for the ROR-SCDA game with $T = 1$ as follows:

- Rand generates uniform random bits always.

- Rev is the identity function.

- Query gets pk and runs $\mathcal{B}_1$ on pk; $\mathcal{B}_1$ outputs a message m and some state state. Query makes a single RoR query where Mesg outputs the fixed message m. Query outputs the resulting ciphertext C as well as state.

- Dist outputs $\mathcal{B}_2(C, \mathsf{state})$.

Since there is no communication between Mesg and Rand, it holds that $c = 0$. Therefore $\mathrm{H}_\infty(\mathbf{pk}) - \min(p,c) = \mathrm{H}_\infty(\mathbf{pk})$. Moreover, the randomness output by Rand is a ℓ-source and the pair (m,r) is a ℓ-source conditioned on ρ (there is no hint ξ). Therefore, the adversary $\mathcal{A}$ satisfies the attack profile $(\mathrm{H}_\infty(\mathbf{pk}), \ell, \ell) = \Pi$.

Finally, the advantage of $\mathcal{A}$ in the ROR-SCDA game is identical to the advantage of $\mathcal{B}$ in the IND-CPA game. ■

4.2 (Weak) CDA Security

In a setting where randomness is bad, the previous notion is no longer achievable. In [BBN+09], Bellare, Brakerski, Naor, Ristenpart, Segev, Shacham and Yilek asked that security is guaranteed as long as the joint distribution of the message and randomness has sufficiently high min-entropy and called this notion *Indistinguishability under a Chosen Distribution Attack* (CDA).

We work again in the "real-or-random" model. The adversary is given access to an oracle RoR that is parametrized by pk, $\mathsf{mode} \in \{\mathsf{real}, \mathsf{random}\}$, and a sampling algorithm $\mathcal{MR}$. Then we define $\mathsf{RoR}_{\mathrm{pk}}(\mathcal{MR}; \mathsf{mode})$ as follows: $(\mathbf{m}, \mathbf{r})$ are distributions over vectors of length $T = \mathrm{poly}(k)$ sampled according to $\mathcal{MR}$. If $\mathsf{mode} = \mathsf{real}$ then output $(\mathsf{Enc}_{\mathrm{pk}}(m_1; r_1), \ldots, \mathsf{Enc}_{\mathrm{pk}}(m_T; r_T))$. Otherwise, sample $(m'_i, r'_i)_{i \in [T]}$ uniformly and independently and output $(\mathsf{Enc}_{\mathrm{pk}}(m'_1; r'_1), \ldots, \mathsf{Enc}_{\mathrm{pk}}(m'_T; r'_T))$.

Definition 4.3. Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption system with security parameter k. An adversary in the ROR-WCDA game is a pair of algorithms $\mathcal{A} = (\mathsf{Query}, \mathsf{Dist})$. The game is defined as follows for $\mathsf{mode} \in \{\mathsf{real}, \mathsf{random}\}$:

$\mathsf{ROR\text{-}WCDA}_{\mathcal{A}}^{\mathsf{mode}}(k)$ game

$(\mathrm{pk}, \mathrm{sk}) \overset{R}{\leftarrow} \mathsf{Gen}(1^k)$

$\tau = \mathsf{Query}_k^{\mathsf{RoR}_{\mathrm{pk}}(\cdot;\mathsf{mode})}(1^k)$

Output $\mathsf{Dist}(\tau, \mathrm{pk})$

$$\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ROR\text{-}WCDA}}(k) = |\Pr[\mathsf{ROR\text{-}WCDA}^{\mathsf{real}}_{\mathcal{A}}(k) = 1] - \Pr[\mathsf{ROR\text{-}WCDA}^{\mathsf{random}}_{\mathcal{A}}(k) = 1]|$$

Definition 4.4 (ROR-WCDA security)**.** A scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is $(\kappa(\cdot), \varepsilon(\cdot,\cdot))$-ROR-WCDA-(block-wise)-secure if $\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ROR\text{-}WCDA}}(k) \leq \varepsilon(k, s)$ for all adversaries $(\mathcal{A}_k)_{k \in \mathbb{N}}$ consisting of $\mathcal{A}_k = (\mathsf{Query}_k, \mathsf{Dist}_k)$ where each algorithm is a circuit of size at most s and for any $\mathcal{MR}$ queried by Query_k, the corresponding distribution $(\mathbf{m}, \mathbf{r})$ is a (block-wise) $\kappa(k)$-source and for every m in the support, all the messages in m are distinct.

Lemma 4.5. *Let $s, \kappa : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \times \mathbb{N} \to [0,1]$. Let $\mathcal{E} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption system with security parameter k, message space $\mathcal{M} = \{0,1\}^{n(k)}$ and randomness space $\mathcal{R} = \{0,1\}^{\ell(k)}$. If $\mathcal{E}$ is $(\Pi(\cdot), 0, \varepsilon(\cdot,\cdot))$-ROR-SCDA-(block-wise)-secure with $\Pi(k) = (\mathrm{H}_\infty(\mathbf{pk}), \ell(k), \kappa(k))$ then $\mathcal{E}$ is $(\kappa(\cdot), \varepsilon(\cdot,\cdot))$-ROR-WCDA-(block-wise)-secure.*

Proof. Assume for contradiction that there exists an adversary $\mathcal{B} = (\mathsf{Query}', \mathsf{Dist}')$ that ε-breaks the ROR-WCDA-security of $\mathcal{E}$. Suppose that Query' makes queries $\mathcal{MR}$ that sample vectors of length $T = \mathrm{poly}(k)$.

We will construct an adversary $\mathcal{A} = (\mathsf{Rand}, \mathsf{Rev}, \mathsf{Query}, \mathsf{Dist})$ for the ROR-SCDA game with the same T as Query' as follows:

- Rand receives input r from Mesg and outputs r.
- Rev is a constant function, revealing nothing about the public key.
- Query runs Query$'$ and for each query $\mathcal{MR}$ made by Query$'$, Query constructs Mesg that samples from $\mathcal{MR}$ to produce (m, r). Mesg outputs m and sends r to Rand. Query forwards the response of the oracle back to Query$'$ and continues. When Query$'$ is done, Query outputs whatever Query$'$ outputs.
- Dist runs Dist$'$ and outputs whatever it outputs.

Observe that $p = |\rho| = 0$ since Rev reveals nothing. Therefore $\mathrm{H}_\infty(\mathbf{pk}) - \min(p(k), c(k)) = \mathrm{H}_\infty(\mathbf{pk})$. Moreover, since $p = 0$, the second entropy condition of Definition 3.2 is vacuously satisfied. Finally, the distribution of message/randomness output by Mesg and Rand is the distribution output by Dist and is therefore a $\kappa(k)$-(block-wise)-source. Therefore, the adversary $\mathcal{A}$ satisfies the attack profile $(\mathrm{H}_\infty(\mathbf{pk}), \ell(k), \ell(k)) = \Pi(k)$.

Finally, the advantage of $\mathcal{A}$ in the ROR-SCDA game is identical to the advantage of $\mathcal{B}$ in the ROR-WCDA game. ■

4.3 Randomness-Dependent Message Security

In [BCPT13], Birell, Chung, Pass and Telang formalized *randomness-dependent message security* of encryption schemes. In this setting, the message to be encrypted may be selected as a function (chosen by the adversary) of the randomness used to encrypt this particular message. They proposed the following notion of bounded strong randomness-dependent message BSRDM) security:

Definition 4.6. For every $s, a, q : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \times \mathbb{N} \to [0, 1]$, a public-key encryption scheme $\mathcal{E} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is a-bounded q-strong ε-BSRDM-CPA-secure if for every adversary $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$ a triple of a circuits of size at most $s(k)$, the ensembles $\{\mathsf{BSRDM\text{-}CPA}^{\mathcal{E}}_{\mathsf{real}}(A, k)\}_{k \in \mathbb{N}}$ and $\{\mathsf{BSRDM\text{-}CPA}^{\mathcal{E}}_{\mathsf{random}}(A, k)\}_{k \in \mathbb{N}}$ are $\varepsilon(k, t(k))$-close where

$$\begin{array}{rcl} \mathsf{BSRDM\text{-}CPA}^{\mathcal{E}}_{\mathsf{mode}}(A, k) & := & (\mathrm{pk}, \mathrm{sk}) \xleftarrow{R} \mathsf{Gen}(1^k) \\ & & (f, \mathsf{state}_1) \xleftarrow{R} \mathcal{A}_1(1^k, \mathrm{pk}) \\ & & r \xleftarrow{R} \{0,1\}^{\ell(k)} \\ & & (m, \mathsf{state}_2) \xleftarrow{R} \mathcal{A}_2(f(r), \mathsf{state}_1) \\ & & \text{if mode} = \mathsf{real}, C \leftarrow \mathsf{Enc}_{\mathrm{pk}}(m, r) \\ & & \text{if mode} = \mathsf{random}, m' \xleftarrow{R} \mathcal{M}, C \leftarrow \mathsf{Enc}_{\mathrm{pk}}(m', r) \\ & & b^* \xleftarrow{R} \mathcal{A}_3(C, \mathsf{state}_2) \\ & & \text{Output } b^* \end{array}$$

Here, $f : \{0,1\}^{\ell(k)} \to \{0,1\}^{q(k)}$ is a function computed by a circuit of size at most $a(k)$.

Lemma 4.7. *Let $s, a, q : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \times \mathbb{N} \to [0, 1]$. Let $\mathcal{E} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption system with security parameter k, message space $\mathcal{M} = \{0,1\}^{n(k)}$ and random space $\mathcal{R} = \{0,1\}^{\ell(k)}$. If $\mathcal{E}$ is $(\Pi(\cdot), \eta(\cdot), \varepsilon(\cdot,\cdot))$-ROR-SCDA-(block-wise)-secure with $\Pi(k) = (\mathrm{H}_\infty(\mathbf{pk}) - a - q, \ell - q - \log(1/\eta), \ell - q - \log(1/\eta))$ then $\mathcal{E}$ is a-bounded q-strong $\varepsilon(\cdot,\cdot)$-BSRDM-CPA-secure.*

Proof. Assume for contradiction that there exists an adversary $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3)$ that ε-breaks the $\mathsf{BSRDM\text{-}CPA}$-security of $\mathcal{E}$ with leakage functions $f : \{0,1\}^\ell \to \{0,1\}^q$ computed by a circuit of size at most $a(k)$.

We will construct an adversary $\mathcal{A} = (\mathsf{Rand}, \mathsf{Rev}, \mathsf{Query}, \mathsf{Dist})$ with $T = 1$ in the $\mathsf{ROR\text{-}SCDA}$ game as follows:

- Rand expects to get a description of a circuit f. It samples r uniformly, sends $f(r)$ back to Mesg, and outputs r.
- Rev is the identity function.
- Query receives pk and runs $\mathcal{B}_1$ on pk; $\mathcal{B}_1$ outputs a leakage function f and some state state_1. Query makes a RoR query Mesg, which contains a description of $\mathcal{B}_2, f, \mathsf{state}_1$ and does the following:

  1. Mesg sends f to Rand and gets back $f(r)$.
  2. Mesg then runs $\mathcal{B}_2(f(r), \mathsf{state}_1)$ which produces a message m and a state state_2.
  3. Mesg outputs m and $\xi = \mathsf{state}_2$.

  Query receives from the RoR oracle a ciphertext C and ξ.
- Dist runs $\mathcal{B}_3(C, \xi)$ and outputs whatever it outputs.

Observe that the communication between $\mathsf{Mesg}, \mathsf{Rand}$ is $c(k) = a(k) + q(k)$ (the size of the description of f and the size of its output). Observe that $p = |\mathrm{pk}|$ and therefore $\mathrm{H}_\infty(\mathbf{pk}) - \min(p(k), c(k)) = \mathrm{H}_\infty(\mathbf{pk}) - a(k) - q(k)$. Moreover, because f shrinks from ℓ to q bits, with probability $1 - \eta$ it holds that $\mathbf{r}_\sigma$ is a $\ell - q - \log(1/\eta)$-source, and with probability $1 - \eta$ the pair (m, r) is a $\ell - q - \log(1/\eta)$-source conditioned on ρ, ξ. Therefore, the adversary $\mathcal{A}$ satisfies the attack profile $(\mathrm{H}_\infty(\mathbf{pk}) - a - q, \ell - q - \log(1/\eta), \ell - q - \log(1/\eta)) = \Pi(k)$ with probability $1 - \eta$.

Finally, the advantage of $\mathcal{A}$ in the $\mathsf{ROR\text{-}SCDA}$ game is identical to the advantage of $\mathcal{B}$ in the $\mathsf{BSRDM\text{-}CPA}$ game. ■

4.4 RSV Security

We use the same RoR oracle as in Section 4.2. We consider only the single-query case as this is without loss of generality for the RSV notion of adaptive security [RSV13a].

Definition 4.8. Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption system with security parameter k. A p-bounded adversary in the ROR-RSV game is a pair of algorithms $\mathcal{A} = (\mathsf{Query}, \mathsf{Dist})$ and an associated family of distributions $\mathcal{F}$ of size $|\mathcal{F}| \le 2^p$ such that the single query $\mathcal{MR}$ made by Query satisfies $\mathcal{MR} \in \mathcal{F}$. The game is defined as follows for $\mathsf{mode} \in \{\mathsf{real}, \mathsf{random}\}$:

$$\boxed{\begin{array}{c} \mathsf{ROR\text{-}RSV}^{\mathsf{mode}}_{\mathcal{A}}(k) \text{ game} \\ \hline (\mathrm{pk}, \mathrm{sk}) \xleftarrow{R} \mathsf{Gen}(1^k) \\ \tau = \mathsf{Query}_k^{\mathsf{RoR}_{\mathrm{pk}}(\cdot;\mathsf{mode})}(\mathrm{pk}) \\ \text{Output } \mathsf{Dist}(\tau, \mathrm{pk}) \end{array}}$$

$$\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ROR\text{-}RSV}}(k) = |\Pr[\mathsf{ROR\text{-}RSV}^{\mathsf{real}}_{\mathcal{A}}(k) = 1] - \Pr[\mathsf{ROR\text{-}RSV}^{\mathsf{random}}_{\mathcal{A}}(k) = 1]|$$

Definition 4.9 (ROR-RSV security)**.** A scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is $(\kappa(\cdot), \varepsilon(\cdot,\cdot))$-ROR-RSV-(block-wise)-secure against p-bounded adversaries if $\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ROR\text{-}RSV}}(k) \leq \varepsilon(k, s)$ for all p-bounded adversaries $(\mathcal{A}_k)_{k \in \mathbb{N}}$ consisting of $\mathcal{A}_k = (\mathsf{Query}_k, \mathsf{Dist}_k)$ where each algorithm is a circuit of size at most s, and Query_k makes a single query whose corresponding distribution $(\mathbf{m}, \mathbf{r})$ is a (block-wise) $\kappa(k)$-source and for every m in the support, all the messages in m are distinct.

Remark 4.10. We comment that in fact the notion of "adaptivity" used in [RSV13a] is different than the notion of adaptive security in the WCDA model or hedged encryption models of [BBN$^+$09]. Although the RSV notion allows the adversary to change its query distribution in each query based on the answers to previous queries, all queries must come from the *same* family of distributions $\mathcal{F}$ of size $|\mathcal{F}| \leq 2^p$. This greatly limits how truly adaptive the reduction may be. In particular, adaptive security in the RSV *does not* imply adaptive security in the WCDA model or hedged encryption model. In order to do so, one would need to allow each adaptive query to be from a *different* family of distributions of size 2^p. Generalizing the definition in this direction (namely by bounding the size of the family of distributions that each query comes from but allowing different families for each query) is cumbersome, as it raises the question of how the definition of the family of distributions may depend on the RoR answers previously seen by the adversary. Therefore, since our SCDA definition subsumes both the adaptive WCDA / hedged models as well as the RSV notion of security, as a nice side-effect it also offers a clean alternative generalization of the RSV security definition to be "truly" adaptive.

Lemma 4.11. *Let $s, \kappa : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \times \mathbb{N} \to [0, 1]$. Let $\mathcal{E} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a public-key encryption system with security parameter k, message space $\mathcal{M} = \{0,1\}^{n(k)}$ and randomness space $\mathcal{R} = \{0,1\}^{\ell(k)}$. If $\mathcal{E}$ is $(\Pi(\cdot), 0, \varepsilon(\cdot,\cdot))$-ROR-SCDA-(block-wise)-secure with $\Pi(k) = (\mathrm{H}_\infty(\mathbf{pk}) - p(k), \ell(k), \kappa(k))$ then $\mathcal{E}$ is $(\kappa(\cdot), \varepsilon(\cdot,\cdot))$-ROR-RSV-(block-wise)-secure against $p(k)$-bounded adversaries.*

Proof. Assume for contradiction that there exists a $p(k)$-bounded adversary $\mathcal{B} = (\mathsf{Query}', \mathsf{Dist}')$ that ε-breaks the ROR-RSV-security of $\mathcal{E}$. Suppose that Query' makes queries $\mathcal{MR}$ that sample vectors of length $T = \mathrm{poly}(k)$.

We will construct an adversary $\mathcal{A} = (\mathsf{Rand}, \mathsf{Rev}, \mathsf{Query}, \mathsf{Dist})$ for the ROR-SCDA game with the same T as Query' as follows:

- Rand receives input r from Mesg and outputs r.
- Rev runs Query' and outputs the choice of $\mathcal{MR}$ to query (this requires only p bits to describe).
- Query queries RoR on the $\mathcal{MR}$ that is computed by Rev. It forwards this response to Query' and outputs whatever Query' outputs.
- Dist runs Dist' and outputs whatever it outputs.

Observe that $p = |\rho|$ is the same as the p-bound on the ROR-RSV adversary. Therefore $\mathrm{H}_\infty(\mathbf{pk}) - \min(p(k), c(k)) \geq \mathrm{H}_\infty(\mathbf{pk}) - p(k)$. Moreover, the second entropy condition of Definition 3.2 is vacuously satisfied. Finally, the distribution of message/randomness output by Mesg and Rand is the distribution output by Dist and is therefore a $\kappa(k)$-(block-wise)-source. Therefore, the adversary $\mathcal{A}$ satisfies the attack profile $(\mathrm{H}_\infty(\mathbf{pk}) - p, \ell(k), \kappa(k)) = \Pi(k)$.

Finally, the advantage of $\mathcal{A}$ in the ROR-SCDA game is identical to the advantage of $\mathcal{B}$ in the ROR-RSV game. ■

5 Constructions

5.1 Standard model

We now show that, using the high moment Crooked Leftover Hash Lemma of [RSV13a] and lossy trapdoor functions, we can construct a ROR-SCDA-blockwise-secure scheme in the standard model. The restriction to block-sources is standard in deterministic encryption and security against chosen distribution attacks (*e.g.* [BFOR08, BFO08, BBN+09, BS11, RSV13a]) and was recently shown by Wichs [Wic13] to be inherent to our techniques.

5.1.1 Main tool

We start by giving our main technical theorem, which is a generalization of Theorem 4.6 from [RSV13a].

Theorem 5.1. *For any $\varepsilon, \eta \in (0,1)$, for any $(n+\ell, a)$-lossy function $f : \{0,1\}^{n+\ell} \to \{0,1\}^{n'}$, and for $\boldsymbol{\pi}$ a t-wise δ-dependent permutation over $\{0,1\}^{n+\ell}$ where $t = b + n + a + \log(T/\varepsilon) + \log(T/\eta) + 1$ and $\delta \leq 2^{-(n+\ell)t}$, the following holds:*
Let $\mathcal{R} = \{\mathbf{r}\}$ be a family of κ block-wise-sources of size $|\mathcal{R}| \leq 2^b$ where $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_T)$ and each $\mathbf{r}_i$ is over $\{0,1\}^\ell$, and where

$$\kappa \geq a + 2\log(1/\varepsilon) + 2\log T + 2\log t + \Theta(1)$$

Then with probability at least $1-\eta$ over the choice of $\pi \in \boldsymbol{\pi}$, it holds that for all $\mathbf{r} \in \mathcal{R}$ and all $m_1, \ldots, m_T \in \{0,1\}^n$ that:

$$\Delta\left((f(\pi(m_1\mathbf{r}_1)), \ldots, f(\pi(m_T\mathbf{r}_T))), f(\mathbf{u}_{n+\ell}^{(1)}), \ldots, f(\mathbf{u}_{n+\ell}^{(T)})\right) \leq \varepsilon$$

where the $\mathbf{u}_{n+\ell}^{(i)}$ are independant uniform strings over $\{0,1\}^{n+\ell}$.

We first remark that a naive application of Theorem 4.6 of [RSV13a] to the hypotheses of Theorem 5.1 would result in a useless bound, since the family of distributions is of size 2^{nT+b} which is much larger than what t-wise δ-dependent permutations can handle for the $t \ll nT$. We are nevertheless able to prove our result because our family of distributions, while large, has a very specific form, and we can apply the analysis of [RSV13a] more carefully to take advantage of this special form.

To prove Theorem 5.1 it is useful to define the following hybrids for each $i \in [T]$: (similar to those defined in the proof of Theorem 4.6 of [RSV13a]):

$$H_i(m, \mathbf{r}) = (m_1\mathbf{r}_1, \ldots, m_{i-1}\mathbf{r}_{i-1}, f(\mathbf{u}_{n+\ell}^{(i)}), \ldots, f(\mathbf{u}_{n+\ell}^{(T)})) \tag{5.1}$$
$$G_i(m, \mathbf{r}) = (m_1\mathbf{r}_1, \ldots, m_{i-1}\mathbf{r}_{i-1}, f(\pi(m_i\mathbf{r}_i)), \ldots, f(\pi(m_T\mathbf{r}_T))) \tag{5.2}$$

Let $H_i(m, \mathbf{r})_{[i]}$ (respectively $G_i(m, \mathbf{r})_{[i]}$) denote the first i components of the vector.

The main ingredient in the proof of Theorem 5.1 is the following lemma:

Lemma 5.2. *Given the hypotheses of Theorem 5.1, for any $i \in [T]$ it holds that*

$$\Pr[\exists m \in \{0,1\}^{in}, \mathbf{r} \in \mathcal{R} \ s.t. \ \Delta(H_i(m, \mathbf{r})_{[i]}, G_i(m, \mathbf{r})_{[i]}) > \varepsilon/T] < \gamma/T$$

Proof. Observe that the first $i-1$ components of both distributions are identical, and also that in both cases the i'th component is independent of $m_1, \ldots, m_{i-1}$. Therefore we have that:

$$\begin{aligned}
&\Pr[\exists m \in \{0,1\}^{in}, \mathbf{r} \in \mathcal{R} \; s.t. \; \Delta(H_i(m,\mathbf{r})_{[i]}, G_i(m,\mathbf{r})_{[i]}) > \varepsilon/T] \\
&\quad = \Pr[\exists m_i \in \{0,1\}^n, \mathbf{r} \in \mathcal{R} \; s.t. \\
&\qquad\qquad \Delta((\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, f(\pi(m_i\mathbf{r}_i))), (\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, f(\mathbf{u}_{n+\ell}))) > \varepsilon/T] \\
&\quad \le \sum_{m_i \in \{0,1\}^n, \mathbf{r} \in \mathcal{R}} \Pr[\Delta((\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, f(\pi(m_i\mathbf{r}_i))), (\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, f(\mathbf{u}_{n+\ell}))) > \varepsilon/T]
\end{aligned}$$

Now applying Lemma 2.6, we see that each term in the sum is bounded by $2^{-n-b}\gamma/T$, and taking the sum we get a bound of γ/T. ■

■

Proof of Theorem 5.1. By definition, proving the theorem is equivalent to proving that

$$\Pr_{\pi \leftarrow_{\mathrm{R}} \boldsymbol{\pi}}[\exists m \in \{0,1\}^{nT}, \mathbf{r} \in \mathcal{R} \; s.t. \; \Delta(H_1(m,\mathbf{r}), G_1(m,\mathbf{r})) > \varepsilon] \le \eta$$

Clearly we cannot take a union bound over all the $m, \mathbf{r}$ since there are too many possibilities of m for us to handle. However we can prove that for each intermediate hybrid we need to take only a union bound over a single m_i and this will allow us to prove the theorem.

Our main claim is the following, which we prove by induction starting at $i = T$.

Claim 5.3. *For each $i \in [T]$, it holds that*

$$\Pr_{\pi \leftarrow_R \boldsymbol{\pi}}[\exists m \in \{0,1\}^{nT}, \mathbf{r} \in \mathcal{R} \; s.t. \; \Delta(H_i(m,\mathbf{r}), G_i(m,\mathbf{r})) > \tfrac{\varepsilon(T-i+1)}{T}] \le \tfrac{\eta(T-i+1)}{T}$$

The base case $i = T$ is just Lemma 5.2 for the setting of $i = T$. The inductive case is proved by splitting up the expression into two parts:

$$\begin{aligned}
&\Pr_{\pi \leftarrow_{\mathrm{R}} \boldsymbol{\pi}}[\exists m \in \{0,1\}^{nT}, \mathbf{r} \in \mathcal{R} \; s.t. \; \Delta(H_i(m,\mathbf{r}), G_i(m,\mathbf{r})) > \tfrac{\varepsilon(T-i+1)}{T}] \\
&\quad = \Pr[\exists m \in \{0,1\}^{in}, \mathbf{r} \in \mathcal{R} \; s.t. \\
&\qquad\qquad \Delta(((m\mathbf{r})_{[i-1]}, f(\pi(m_i\mathbf{r}_i)), f(\mathbf{u}^{(i+1)}_{n+\ell}), \ldots f(\mathbf{u}^{(T)}_{n+\ell})), \\
&\qquad\qquad\quad ((m\mathbf{r})_{[i-1]}, f(\mathbf{u}^{(i)}_{n+\ell}), \ldots, f(\mathbf{u}^{(T)}_{n+\ell}))) > \tfrac{\varepsilon}{T}] \\
&\qquad + \Pr[\exists m \in \{0,1\}^{nT}, \mathbf{r} \in \mathcal{R} \; s.t. \\
&\qquad\qquad \Delta(((m\mathbf{r})_{[i-1]}, f(\pi(m_i\mathbf{r}_i)), f(\mathbf{u}^{(i+1)}_{n+\ell}), \ldots f(\mathbf{u}^{(T)}_{n+\ell})), \\
&\qquad\qquad\quad ((m\mathbf{r})_{[i-1]}, f(\pi(m_i\mathbf{r}_i)), \ldots, f(\pi(m_T\mathbf{r}_T)))) > \tfrac{\varepsilon(T-i)}{T}] \qquad (5.3) \\
&\quad \le \tfrac{\eta}{T} + \tfrac{\eta(T-i)}{T}
\end{aligned}$$

The first term of Equation 5.3 is bounded by Lemma 5.2 because the coordinates after the i'th coordinate are identical in both distributions and independent of the previous coordinates. The second term is by the inductive hypothesis and the fact that the distance at the i'th coordinate cannot increase by applying the function $f(\pi(\cdot))$ to the same random variable $m_i\mathbf{r}_i$ in the i'th coordinate in both distributions. ■

Fix any polynomial $a(k), b(k)$. Fix any polynomials $n(k)$ (message length) and $\ell(k)$ (randomness length) such that $\ell(k) = \omega(a(k))$ and also $a(k) = (n(k) + \ell(k))^{\Omega(1)}$. Fix a family of $(n + \ell, a)$-lossy trapdoor functions. Fix any $T = k^{\omega(1)}$ and set $t = b + n + a + 2\log T + \omega(\log k)$. We define the following public-key encryption:

Key generation : $\mathsf{Gen}(1^k)$ runs the injective generation function for a $(n + \ell, a)$-lossy trapdoor function family to generate (f, f^{-1}). $\mathsf{Gen}(1^k)$ also samples $\pi \leftarrow_{\mathrm{R}} \boldsymbol{\pi}$ from a t-wise δ-dependent permutation over $\{0,1\}^{n+\ell}$. The public key is $\mathrm{pk} = (f, \pi)$ and the private key is $\mathrm{sk} = f^{-1}$.

Encryption : $\mathsf{Enc}_{\mathrm{pk}}(m; r) = f(\pi(m\|r))$ where $r \in \{0,1\}^{\ell(k)}$.

Decryption : $\mathsf{Dec}_{\mathrm{sk}}(c) = \pi^{-1}(f^{-1}(c))_{[n]}$.

Algorithm 5.4. ROR-SCDA encryption in standard model

5.1.2 Definition and security of our scheme

Our scheme is essentially the same as the scheme of [RSV13a] with a longer key size, which seems necessary to achieve our stronger notion of security. Our scheme is defined in Algorithm 5.4.

Theorem 5.5. *Fix any $\kappa(k) \geq a + 2\log t + \omega(\log k)$. The encryption system in Algorithm 5.4 is Π-ROR-SCDA-blockwise-secure for $\Pi = (\mathrm{H}_\infty(\mathbf{pk}) - b, \kappa, \kappa)$.*

Proof. We in fact prove that the scheme satisfies a somewhat stronger notion of security. Using this stronger notion aids in our proof as it will allow us to reduce from the multi-query adaptive case to the single-query non-adaptive case.

Intuitively, we will require the scheme to be secure even if the adversary sees f for free (where $\mathrm{pk} = (f, \pi)$) and even if we require that the ciphertexts are indistinguishable from random ciphertexts encrypted by (f, π') for an independent random π' (rather than being encrypted by the same (f, π). This stronger requirement is similar to anonymity (or key-privacy) [BBDP01] and is necessary to show that no additional information is leaked about the permutation π in the adaptive queries.

Formally, we define the oracle $\mathsf{RoR}'_{(f,\pi)}$ as follows:

1. $\mathsf{RoR}'_{(f,\pi)}(\mathsf{Mesg}; \mathsf{Rand}_k, \mathsf{real}, T)$ behaves identically to $\mathsf{RoR}_{(f,\pi)}(\mathsf{Mesg}, \mathsf{Rand}_k, \mathsf{real}, T)$.
2. $\mathsf{RoR}'_{(f,\pi)}(\cdot; \mathsf{Rand}_k, \mathsf{random}, T)$ behaves identically to $\mathsf{RoR}_{(f,\pi)}(\cdot, \mathsf{Rand}_k, \mathsf{random}, T)$ (*i.e.* encrypting uniformly chosen messages and randomness) *except* that the encryption is performed using (f, π') for an independently and randomly chosen π' rather than using (f, π).

Definition 5.6 (ROR-SCDA$'$ game)**.** An adversary in the ROR-SCDA game is a sequence of integers $T_k = \mathrm{poly}(k)$ and tuples of algorithms $\mathcal{A}_k = (\mathsf{Rand}_k, \mathsf{Rev}_k, \mathsf{Query}_k, \mathsf{Dist}_k)$. The game is defined as follows for $\mathsf{mode} \in \{\mathsf{real}, \mathsf{random}\}$:

$(\textsf{ROR-SCDA}')^{\textsf{mode}}_{\mathcal{A}}(k)$ game

$$(\mathrm{pk} = (f, \pi), \mathrm{sk} = f^{-1}) \overset{R}{\leftarrow} \textsf{Gen}(1^k)$$
$$\tau = \textsf{Query}_k^{\textsf{RoR}'_{(f,\pi)}(\cdot;\textsf{Rand}_k,\textsf{mode},T_k)}(f, \textsf{Rev}_k(\mathrm{pk}))$$
$$\text{Output } \textsf{Dist}_k(\tau, \mathrm{pk})$$

$$\mathrm{Adv}^{\mathcal{A}}_{\textsf{ROR-SCDA}'}(k) = |\Pr[(\textsf{ROR-SCDA}')^{\textsf{real}}_{\mathcal{A}}(k) = 1] - \Pr[(\textsf{ROR-SCDA}')^{\textsf{random}}_{\mathcal{A}}(k) = 1]|$$

First observe that for every $\mathcal{A}$, there exists $\mathcal{A}'$ of the same computational complexity such that

$$\mathrm{Adv}^{\mathcal{A}}_{\textsf{ROR-SCDA}}(k) \leq \mathrm{Adv}^{\mathcal{A}}_{\textsf{ROR-SCDA}'}(k) + \mathrm{Adv}^{\mathcal{A}'}_{\textsf{ROR-SCDA}'}(k) \tag{5.4}$$

The adversary $\mathcal{A}'$ emulates $\mathcal{A}$ and for each RoR query that $\mathcal{A}$ makes, calls $\textsf{RoR}'$ with *uniform* message/randomness distributions and forwards the answers back to the emulated $\mathcal{A}$. $\mathcal{A}'$ then runs the same distinguisher as $\mathcal{A}$. The above inequality holds by the triangle inequality, since $\mathcal{A}$ in game $\textsf{ROR-SCDA}'$ gets oracle answers that are either real encryptions of the queried message distributions or else random messages encrypted with (f, π'), while $\mathcal{A}'$ is the same as running $\mathcal{A}$ with oracle answers that are either random messages encrypted with (f, π') or random messages encrypted with the true (f, π).

Next observe that one can reduce the case of many-query $\textsf{ROR-SCDA}'$ to the single-query case: consider any adversary $\mathcal{A}$ that makes at most $q = \mathrm{poly}(k)$ queries to the $\textsf{RoR}'$ oracle. Consider single-query adversaries $\mathcal{A}_1, \ldots, \mathcal{A}_q$ where for each $\mathcal{A}_i$, all queries $< i$ are answered according to $\textsf{random}$ mode and all queries $> i$ are answered according to $\textsf{real}$ mode, and the i'th query is sent to the actual $\textsf{RoR}'$ oracle. Observe that the queries $< i$ can be answered by $\mathcal{A}_i$ because we give the query algorithm f for free and so it can sample messages and π' on its own to compute random encryptions. Furthermore, observe that $\mathcal{A}_i$ can answer queries $> i$ because after its last/only query to the actual $\textsf{RoR}'$ oracle we may push the rest of the execution of $\mathcal{A}$ into the distinguisher, where the adversary has full access to $\mathrm{pk} = (f, \pi)$ and so it can compute real encryptions by itself. Thus it holds that

$$\max_{\text{efficient } \mathcal{A} \text{ making } q \text{ queries}} \mathrm{Adv}^{\mathcal{A}}_{\textsf{ROR-SCDA}'}(k) \quad \leq \quad q \cdot \left(\max_{\text{efficient } \mathcal{A} \text{ making single query}} \mathrm{Adv}^{\mathcal{A}}_{\textsf{ROR-SCDA}'}(k) \right) \tag{5.5}$$

Therefore from Equation 5.4 and Equation 5.5, to prove the ROR-SCDA-security of the scheme it suffices to show that the advantage of a single-query adversary in the $\textsf{ROR-SCDA}'$ game is negligible.

The remainder of the proof breaks into two cases depending on the profile of the attacker. Our attacker is a tuple of algorithms $\mathcal{A}_k = (\textsf{Rand}_k, \textsf{Rev}_k, \textsf{Query}_k, \textsf{Dist}_k)$, and we have the guarantee that $\textsf{Query}_k$ makes exactly one query to the RoR oracle. In fact, we can also assume without loss of generality that $\textsf{Query}_k$ is deterministic, since it can push all of its randomness into its single query $\textsf{Mesg}$. Also recall that the input to $\textsf{Query}_k$ is a string ρ of length p and that σ denotes the communication between $\textsf{Mesg}_k$ and $\textsf{Rand}_k$, and $c = |\sigma|$. We distinguish the following two cases.

If $p \leq b$: in this case the input to $\textsf{Query}_k$ can be one of at most 2^b values, and therefore the output distribution $(\boldsymbol{\xi}, \mathbf{m}, \mathbf{r})$ of the interaction of $\textsf{Mesg}, \textsf{Rand}$ can be one of at most 2^b possible distributions. Furthermore, by the third entropy condition of Definition 3.2, $(\mathbf{m}, \mathbf{r})$ form a κ-source, even when conditioned on the values of ρ, ξ. The theorem follows by considering the following sequence of games: $\mathcal{G}_0$ is just actual $\textsf{ROR-SCDA}'$ security game in $\textsf{real}$ mode.

$\mathcal{G}_1$ is like $\mathcal{G}_0$ except LGen_0 is used to sample a lossy function instead of LGen_1. $\mathcal{G}_0$ and $\mathcal{G}_1$ are indistinguishable because of the indistinguishability of lossy and injective functions sampled from LGen_0 and LGen_1.

$\mathcal{G}_2$ is like $\mathcal{G}_1$ except the RoR oracle is in random mode. $\mathcal{G}_1$ and $\mathcal{G}_2$ are indistinguishable because of Theorem 5.1: in $\mathcal{G}_1$ the distinguisher sees $\mathbf{X}_1 = (\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \mathbf{f}(\boldsymbol{\pi}(\mathbf{m}_1\mathbf{r}_1)), \ldots, \mathbf{f}(\boldsymbol{\pi}(\mathbf{m}_T\mathbf{r}_T)))$ while in $\mathcal{G}_2$ the distinguisher sees $\mathbf{X}_2 = (\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \mathbf{f}(\boldsymbol{\pi}'(\mathbf{u}_{n+\ell}^{(1)})), \ldots, \mathbf{f}(\boldsymbol{\pi}'(\mathbf{u}_{n+\ell}^{(T)})))$. Since $p \leq b$, this means there are at most 2^b distinct possible inputs to $\langle \mathsf{Mesg}_k, \mathsf{Rand}_k \rangle$, $(\mathbf{m}, \mathbf{r})$ can be one of at most 2^b possible distributions.

Furthermore, with overwhelming probability $(\mathbf{m}, \mathbf{r})$ is a κ-block-wise-source, even conditioned on f, π, ξ. This holds because by the third entropy condition of Definition 3.3, $(\mathbf{m}, \mathbf{r})$ is a κ-block-wise-source conditioned on ρ, ξ with overwhelming probability over ξ, and the only dependence that $(\mathbf{f}, \boldsymbol{\pi})$ have on $\mathbf{m}, \mathbf{r}$ is through $\boldsymbol{\rho}$, which is fixed. Therefore we can apply Theorem 5.1 where the $\mathbf{m}$ in the hypothesis of Theorem 5.1 is empty and the $\mathbf{r}$ in its hypothesis is the $(\mathbf{m}, \mathbf{r})$ in our game. We deduce that $\mathbf{X}_1$ has negligible statistical distance to $(\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \mathbf{f}(\mathbf{u}_{n+\ell})^T)$, which is distributed identically to $\mathbf{X}_2$ since $\boldsymbol{\pi}'$ is a permutation.

$\mathcal{G}_3$ is like $\mathcal{G}_2$ except LGen_1 is used to sample an injective function instead of LGen_0. Again this follows by the security of lossy trapdoor functions.

We conclude by noting that $\mathcal{G}_3$ is the $\mathsf{ROR\text{-}SCDA}'$ security game in random mode.

If $p > b$: in this case, the first entropy condition of Definition 3.2 implies that $c \leq b$. Furthermore, the second entropy condition implies that $\mathbf{r}_\sigma$ is a κ-block-wise-source with overwhelming probability over σ. Security follows from the following sequence of games. $\mathcal{G}_0$ is just the actual $\mathsf{ROR\text{-}SCDA}'$ security game in real mode.

$\mathcal{G}_1$ is like $\mathcal{G}_0$ except LGen_0 is used to sample a lossy function instead of LGen_1. $\mathcal{G}_0$ and $\mathcal{G}_1$ are indistinguishable because of the indistinguishability of lossy and injective functions sampled from LGen_0 and LGen_1.

$\mathcal{G}_2$ is like $\mathcal{G}_1$ except the RoR oracle is in random mode. $\mathcal{G}_1$ and $\mathcal{G}_2$ are indistinguishable because of Theorem 5.1: the view of the adversary in $\mathcal{G}_1$ is the tuple $\mathbf{X}_1 = (\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \mathbf{f}(\boldsymbol{\pi}(\mathbf{m}_1\mathbf{r}_1)), \ldots, \mathbf{f}(\boldsymbol{\pi}(\mathbf{m}_T\mathbf{r}_T)))$ while the view of the adversary in $\mathcal{G}_2$ is $\mathbf{X}_2 = (\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \mathbf{f}(\boldsymbol{\pi}'(\mathbf{u}_{n+\ell}^{(1)})), \ldots, \mathbf{f}(\boldsymbol{\pi}'(\mathbf{u}_{n+\ell}^{(T)})))$. Since $\boldsymbol{\pi}'$ is a permutation for each π' in its support, it holds that $\mathbf{X}_2$ is distributed *identically* to $(\mathbf{f}, \boldsymbol{\pi}, \mathbf{f}(\mathbf{u}_{n+\ell}^{(1)}), \ldots, \mathbf{f}(\mathbf{u}_{n+\ell}^{(T)}))$. Suppose now we reveal $\mathbf{m}$ and $\boldsymbol{\sigma}$ to the adversary for free (this can only increase statistical distance, so bounding the distance with this additional revealed information suffices). Since there are at most 2^b possible transcripts σ, $\mathbf{r}_\sigma$ can be one of at most 2^b possible distributions, and due to the second entropy condition of Definition 3.2 and the fact that $\beta = \kappa$ in our attacker profile, it holds that with overwhelming probability over σ, $\mathbf{r}_\sigma$ is a κ-source. Letting $\mathbf{f}(\boldsymbol{\pi}(\mathbf{mr}))$ denote $\mathbf{f}(\boldsymbol{\pi}(\mathbf{m}_1\mathbf{r}_1)) \ldots \mathbf{f}(\boldsymbol{\pi}(\mathbf{m}_T\mathbf{r}_T))$, we want to bound:

$$\Delta((\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \sigma, \mathbf{m}, \mathbf{f}(\boldsymbol{\pi}(\mathbf{m}\mathbf{r}_{\boldsymbol{\sigma}}))), (\mathbf{f}, \boldsymbol{\xi}, \boldsymbol{\pi}, \sigma, \mathbf{m}, \mathbf{f}(\mathbf{u}_{n+\ell})^T)) \tag{5.6}$$

$$\leq \mathbb{E}_{(f,\xi,\pi,\sigma,m) \leftarrow_{\mathrm{R}} (\mathbf{f},\boldsymbol{\xi},\boldsymbol{\pi},\boldsymbol{\sigma},\mathbf{m})} \left[\Delta\left(f(\pi(m\mathbf{r}_\sigma) \mid \boldsymbol{\xi} = \xi), f(\mathbf{u}_{n+\ell})^T\right)\right] \tag{5.7}$$

$$\leq \mathbb{E}_{(f,\pi) \leftarrow_{\mathrm{R}} (\mathbf{f},\boldsymbol{\pi})} \left[\max_{\sigma,m} \Delta\left(f(\pi(m\mathbf{r}_\sigma)), f(\mathbf{u}_{n+\ell})^T\right)\right] \tag{5.8}$$

We may remove the dependence on ξ going from Equation 5.7 to Equation 5.8 because ξ is output by Mesg, and the only way Mesg can influence Rand is through their communication σ, and therefore the only dependence of $\mathbf{r}_\sigma$ on ξ is through σ, which is already fixed. Finally, we can apply Theorem 5.1 and our choice of t large enough to say that for each f, with overwhelming probability over the choice of $\pi \leftarrow_{\mathrm{R}} \boldsymbol{\pi}$, the statistical distance is negligible for all choices of σ, m, and therefore the above expression is negligible.

$\mathcal{G}_3$ is like $\mathcal{G}_2$ except LGen_1 is used to sample an injective function instead of LGen_0. Again this follows by the security of lossy trapdoor functions.

We conclude by noting that $\mathcal{G}_3$ is just the $\mathsf{ROR\text{-}SCDA}'$ security game in random mode.

■

5.2 From random oracle

We construct a scheme secure in the random oracle model. We explicitly only treat the ROR-SCDA case; the ROR-SCDCA case (with a decryption oracle) is entirely analogous.

Although we would like our construction to work starting with any IND-CPA scheme in the standard model, this turns out to be impossible. We will need the starting scheme to be anonymous, *i.e.* ciphertexts do not leak information about the public key.

Let $\mathsf{ARoR}_{\mathrm{pk}}(\mathsf{Mesg};\mathsf{mode})$ for $\mathsf{mode} \in \{\mathsf{real}, \mathsf{anon}\}$ be defined as follows: if $\mathsf{mode} = \mathsf{real}$ then it samples $m \leftarrow_{\mathrm{R}} \mathsf{Mesg}$ and uniform r and outputs $\mathsf{Enc}_{\mathrm{pk}}(m; r)$, else if $\mathsf{mode} = \mathsf{anon}$ it samples $(\mathrm{pk}', sk') \leftarrow_{\mathrm{R}} \mathsf{Gen}(1^k)$ and m', r' uniform and outputs $\mathsf{Enc}_{\mathrm{pk}'}(m'; r')$.

Definition 5.7 (ANON-CPA game)**.** Fix a public-key cryptosystem $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$. The ANON-CPA game is defined as follows for $\mathsf{mode} \in \{\mathsf{real}, \mathsf{anon}\}$ and for an adversary $\mathcal{A} = (\mathsf{Mesg}, \mathsf{Dist})$:

$\mathsf{ANON\text{-}CPA}^{\mathsf{mode}}_{\mathcal{A}}$ game

$(\mathrm{pk}, \mathrm{sk}) \xleftarrow{R} \mathsf{Gen}(1^k)$
$\mathsf{mode} \xleftarrow{R} \{\mathsf{real}, \mathsf{anon}\},\ c = \mathsf{ARoR}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{mode})$
Output $\mathsf{Dist}(\mathrm{pk}, c)$

Define

$$\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ANON\text{-}CPA}}(k) = |\Pr[\mathsf{ANON\text{-}CPA}^{\mathsf{real}}_{\mathcal{A}}(k) = 1] - \Pr[\mathsf{ANON\text{-}CPA}^{\mathsf{anon}}_{\mathcal{A}}(k) = 1]|$$

Definition 5.8 (ANON-CPA security)**.** $(\mathsf{RGen}, \mathsf{REnc}, \mathsf{RDec})$ is a $\varepsilon(k, s)$-ANON-CPA secure scheme if it is $\varepsilon(k, s)$-IND-CPA and in addition for any adversary $\mathcal{A}$ computable in size s it holds that

$$\mathrm{Adv}^{\mathcal{A}}_{\mathsf{ANON\text{-}CPA}}(k) \leq \varepsilon(k, s)$$

We remark that one can construct schemes with ANON-CPA security under standard assumptions [BBDP01].

Definition 5.9. Let $(\mathsf{RGen}, \mathsf{REnc}, \mathsf{RDec})$ be an ANON-CPA secure scheme where for security parameter k, REnc uses $\ell = \ell(k)$ random bits. Let $\mathsf{RO}_\ell : \{0,1\}^* \to \{0,1\}^\ell$ be a random oracle with ℓ-bit outputs. We define $(\mathsf{ROGen}, \mathsf{ROEnc}, \mathsf{RODec})$ where:

- $\mathsf{ROGen} = \mathsf{RGen}$.
- $\mathsf{ROEnc}_{\mathrm{pk}}(m; r) = \mathsf{REnc}_{\mathrm{pk}}(m; \mathsf{RO}_\ell(\mathrm{pk} \parallel m \parallel r))$, where $r \in \{0,1\}^k$.

- $\mathsf{RODec}_{\mathrm{sk}}(c) = \mathsf{RDec}_{\mathrm{sk}}(c)$.

Theorem 5.10. *Suppose that* $(\mathsf{RGen}, \mathsf{REnc}, \mathsf{RDec})$ *is a* $\varepsilon^{\mathsf{REnc}}(k,s)$*-*ANON-CPA*-secure scheme. Then for any function* $\kappa(k)$*, let* $\Pi(k) = (\kappa(k), \kappa(k) - (\mathrm{H}_\infty(\mathbf{pk}) - p(k)), \kappa(k))$*, it holds that the scheme* $(\mathsf{ROGen}, \mathsf{ROEnc}, \mathsf{RODec})$ *defined in Definition 5.9 is* $(\Pi(k), \eta(k), \varepsilon(k,s))$*-*ROR-SCDA*-secure where*

$$\varepsilon(k,s) \le 2s\eta(k) + 3s \cdot \varepsilon^{\mathsf{REnc}}(k,s) + (2s^2+1)2^{-\kappa(k)/2+2}$$

Proof. Fix any adversary $\mathcal{A} = (\mathcal{A})_{k \in \mathbb{N}}$ against $(\mathsf{ROGen}, \mathsf{ROEnc}, \mathsf{RODec})$ in the ROR-SCDA game. Fix k, s and $p = p(k), c = c(k), \kappa = \kappa(k)$. For each of the s queries that Query may make, the probability that the second or third entropy conditions of Definition 3.2 are violated is at most η, and so by a union bound the probability that they are violated in any query is bounded by $2s\eta(k)$. Therefore, in the remainder, we condition on all entropy conditions holding.

Game $\mathcal{G}_0$ is $\mathsf{ROR\text{-}SCDA}_{\mathcal{A}}^{\mathsf{real}}(k)$.

Games $\mathcal{G}_i^{\mathsf{mode}}$ **for** $i \in [s]$ **and** $\mathsf{mode} \in \{\mathsf{real}, \mathsf{anon}\}$. Let us assume without loss of generality that Query_k makes s queries to the RoR oracle (if it makes fewer, just fill the remainder with dummy queries).

Game $\mathcal{G}_i^{\mathsf{real}}$ is like $\mathcal{G}_{i-1}^{\mathsf{anon}}$ (set $\mathcal{G}_0^{\mathsf{anon}} = \mathcal{G}_0$) except that the i'th call to the RoR oracle is answered with an altered oracle $\widetilde{\mathsf{RoR}}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{Rand}_k, \mathsf{real})$. In this oracle, if we let m be the message vector sampled by an interaction of $\mathsf{Mesg}, \mathsf{Rand}_k$, and then the oracle encrypts it using the underlying encryption and *fresh randomness, i.e.* it returns $(\mathsf{REnc}_{\mathrm{pk}}(m_i, r'_i))_{i \in [T]}$ where r' is sampled independently and uniformly from $\{0,1\}^{\ell(k)}$.

For $\mathcal{G}_i^{\mathsf{anon}}$ is just like $\mathcal{G}_i^{\mathsf{real}}$ except that the i'th oracle call to the RoR oracle is answered with T ciphertexts sampled by $\mathsf{ARoR}_{\mathrm{pk}}(\mathbf{u}_n; \mathsf{anon})$, *i.e.* each ciphertext is a fresh encryption of uniformly random m_q, r_q using a completely independently sampled $\mathrm{pk}'_q \xleftarrow{R} \mathsf{RGen}(1^k)$ for $q \in [T]$. (So in game $\mathcal{G}_s^{\mathsf{anon}}$, all of the RoR are answered with encryptions of random m_q, r_q each with independently chosen pk'_q.)

Lemma 5.11. *For each* $i \in [s]$*,* $\mathcal{G}_i^{\mathsf{real}}$ *is* ε'*-indistinguishable from* $\mathcal{G}_{i-1}^{\mathsf{anon}}$ *for*

$$\varepsilon' = (5s^2 + 2) \cdot 2^{-\kappa/2} + 2s \cdot \varepsilon^{\mathsf{REnc}}(k,s)$$

Proof. We claim that the only way $\mathcal{G}_i^{\mathsf{real}}$ and $\mathcal{G}_{i-1}^{\mathsf{anon}}$ can be distinguished is if in $\mathcal{G}_i^{\mathsf{real}}$ the following bad event occurs: if we let Mesg be the j'th query made by Query_k and let (m, r) be the sampled vectors for this query, it holds for some $q \in [T]$, that $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ is queried by Query_k, by Dist_k, or by Mesg or Rand_k during the interaction that produces (m, r). If the bad event does not occur, $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ is always queried exactly once during the entire game and so we may replace it by uniform randomness, which is what is done in $\widetilde{\mathsf{RoR}}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{Rand}_k, \mathsf{real})$.

We know that the number of RO queries made by each of Query_k, Dist_k, Mesg, and Rand_k is at most s. Let us examine each RoR query made by Query_k. For the j'th query, say that Mesg is the query, and let (m, r) be the resulting sample of message/randomness vectors and ξ be the associated hint. Then we argue:

- The probability that for the sampled (m, r) for this query there exists $q \in [T]$ such that $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ was queried by Query_k *prior* to querying Mesg is at most $Ts2^{-\kappa} \le s^2 2^{-\kappa}$. This is by a union bound over q and because Query_k makes at most s queries to RO and because we are guaranteed that $\mathrm{H}_\infty(\mathbf{m}_q, \mathbf{r}_q) \ge \kappa$.

- The probability that on this query, Mesg queries $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ during the sampling interaction is at most $2^{-\kappa/2} + s^2 2^{-\kappa/2}$.

  To deduce this fact, let $\boldsymbol{\rho} = \mathsf{Rev}_k(\mathbf{pk})$ and observe that conditioned on $\boldsymbol{\rho} = \rho$ it holds that $\mathbf{pk}$ is independent of $\boldsymbol{\sigma}, \mathbf{r}$, where σ is the transcript between $\mathsf{Mesg}, \mathsf{Rand}_k$ and r is the randomness output by Rand_k for the j'th query. This is because the only information about $\mathbf{pk}$ that Query_k has before the i'th query in $\mathcal{G}^{\mathsf{anon}}_{i-1}$ is ρ, since all the other RoR queries are answered using fresh public keys.)

  Similarly, for any σ, conditioned on $\boldsymbol{\sigma} = \sigma$ it holds that $\mathbf{r}$ is independent of $\boldsymbol{\rho}, \mathbf{pk}$. Therefore we may write for each $q \in [T]$:

  $$\begin{aligned} \mathrm{H}_\infty(\mathbf{pk}, \mathbf{r}_q \mid (\boldsymbol{\rho}, \boldsymbol{\sigma}) = (\rho, \sigma)) &= \mathrm{H}_\infty(\mathbf{pk} \mid (\boldsymbol{\rho}, \boldsymbol{\sigma}) = (\rho, \sigma)) \\ &\quad + \mathrm{H}_\infty(\mathbf{r}_q \mid (\boldsymbol{\rho}, \boldsymbol{\sigma}) = (\rho, \sigma)) \\ &= \mathrm{H}_\infty(\mathbf{pk} \mid \boldsymbol{\rho} = \rho) + \mathrm{H}_\infty(\mathbf{r}_i \mid \boldsymbol{\sigma} = \sigma) \end{aligned} \tag{5.9}$$

  We may then apply Lemma 2.1 to the first term (setting $\mathbf{x} = \mathbf{pk}$ and $\mathbf{y} = \boldsymbol{\rho}$) to deduce that with probability at least $1 - 2^{-\kappa/2}$ that we pick $\rho \leftarrow_{\mathrm{R}} \boldsymbol{\rho}$ such that

  $$\mathrm{H}_\infty(\mathbf{pk} \mid \boldsymbol{\rho} = \rho) \geq \mathrm{H}_\infty(\mathbf{pk}) - p - \kappa/2$$

  In this case, combining with Equation 5.9 and the second entropy condition of Definition 3.3 with our choice of β, we conclude that

  $$\mathrm{H}_\infty(\mathbf{pk}, \mathbf{r}_q \mid (\boldsymbol{\rho}, \boldsymbol{\sigma}) = (\rho, \sigma)) \geq \kappa/2$$

  Since the only dependence of Mesg on $\mathbf{pk}, \mathbf{r}$ is through $\boldsymbol{\rho}, \boldsymbol{\sigma}$, therefore with probability at least $1 - 2^{-\kappa/2}$ the entropy is high and in this case the probability that Mesg_k makes the bad query is at most $Ts2^{-\kappa/2} \leq s^2 2^{-\kappa/2}$.

- The probability that Rand_k queries $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ is $\leq 2^{-\kappa/2} + s^2 2^{-\kappa/2}$. This is essentially for the same reason as the above point: since $\min(n, c) \leq \mathrm{H}_\infty(\mathbf{pk}) - \kappa$, it holds that with probability at least $1 - 2^{-\kappa/2}$ that $\mathrm{H}_\infty(\mathbf{pk} \mid \boldsymbol{\sigma} = \sigma) \geq \kappa/2$, and in this case the probability that Rand_k makes the bad query is bounded by $s^2 2^{-\kappa/2}$.

- The probability that Query_k queries $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ after querying ρ *and* none of the previous queries were bad is at most $s \cdot \varepsilon^{\mathsf{REnc}}(k, s) + s^2 2^{-\kappa}$. To see this, observe that since none of the previous queries were bad, in fact the encryptions of $m_1, \ldots, m_T$ were performed using true randomness, and so we may replace the oracle's response $(\mathsf{ROEnc}_{\mathrm{pk}}(m_q, r_q))_{q \in [T]}$ with $(\mathsf{REnc}_{\mathrm{pk}}(\mathbf{u}_n, r'_q))$ and this incurs at most a $s\varepsilon^{\mathsf{REnc}}(k, s)$ change in probability. But in this latter case, Query_k has no information about (m, r). By the third entropy condition of Definition 3.2, we know that $(\mathbf{m}, \mathbf{r})$ is a κ-source even conditioned on ρ, ξ, and so it follows that the probability that Query_k makes a bad query is at most $Ts2^{-\kappa} \leq s^2 2^{-\kappa}$.

- For the same reason as in the previous point, the probability that Dist_k queries $\mathsf{RO}(\mathrm{pk} \parallel m_q \parallel r_q)$ after querying ρ *and* none of the previous queries were bad is at most $s \cdot \varepsilon^{\mathsf{REnc}}(k, s) + s^2 2^{-\kappa}$.

Observing that $2^{-\kappa} < 2^{-\kappa/2}$ and collecting all these terms gives the bound on ε'. ■

Claim 5.12. *For each $i \in [s]$, $\mathcal{G}^{\mathsf{real}}_i$ is $s \cdot \varepsilon^{\mathsf{REnc}}(k, s)$-indistinguishable from $\mathcal{G}^{\mathsf{anon}}_i$.*

Proof. This follows from the $\mathsf{ANON\text{-}CPA}$-security of the original scheme: between $\mathcal{G}_i^{\mathsf{real}}$ and $\mathcal{G}_i^{\mathsf{anon}}$ the only difference is that the i'th query is answered either by a real encryption using pk or by an encryption of a uniform message and randomness using a fresh random key. Since the i'th query consists of encryptions of $T \leq s$ messages, the bound follows. ■

Games $\mathcal{G}_i^{\mathsf{mode}}$ for $\mathsf{mode} \in \{\mathsf{unif}, \mathsf{random}\}$ are defined inductively as follows. For $i \in [s]$, $\mathcal{G}_i^{\mathsf{unif}}$ be like $\mathcal{G}_{i+1}^{\mathsf{random}}$ (set $\mathcal{G}_{s+1}^{\mathsf{random}} = \mathcal{G}_s^{\mathsf{anon}}$) except that we answer the i'th RoR query with $\widetilde{\mathsf{RoR}}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{Rand}_k, \mathsf{random})$, which samples random messages and encrypts them using the original scheme with the true pk and with uniform independent randomness.

The game $\mathcal{G}_i^{\mathsf{random}}$ is like $\mathcal{G}_i^{\mathsf{unif}}$ except that the i'th query is answered using the original random oracle $\mathsf{RoR}_{\mathrm{pk}}(\mathsf{Mesg}; \mathsf{Rand}_k, \mathsf{random})$.

Claim 5.13. *For all $i \in [s]$, $\mathcal{G}_i^{\mathsf{unif}}$ is $s \cdot \varepsilon^{\mathsf{REnc}}(k, s)$-indistinguishable from $\mathcal{G}_{i+1}^{\mathsf{random}}$.*

Proof. This follows from the $\mathsf{ANON\text{-}CPA}$-security of the original scheme: between $\mathcal{G}_i^{\mathsf{unif}}$ and $\mathcal{G}_{i+1}^{\mathsf{random}}$ the only difference is that the i'th query is either vector of encryptions of uniform messages by the real key pk or a vector of encryptions of uniform messages using a fresh independent random key. ■

Claim 5.14. *For all $i \in [s]$, $\mathcal{G}_i^{\mathsf{unif}}$ is ε'-indistinguishable from $\mathcal{G}_i^{\mathsf{random}}$.*

Proof. Proved identically to the proof of Lemma 5.11, noting that in the case of uniformly random $\mathbf{m}', \mathbf{r}'$, the entropy conditions we use are trivially satisfied. ■

Finally, it is clear that $\mathcal{G}_1^{\mathsf{random}}$ is just $\mathsf{ROR\text{-}SCDA}_{\mathcal{A}}^{\mathsf{random}}(k)$. Combining all these claims, we get that the distinguishing advantage is

$$\begin{aligned}
\mathrm{Adv}_{\mathsf{ROR\text{-}SCDA}}^{\mathcal{A}}(k) &\leq 2s \cdot \eta(k) + 2s \cdot \varepsilon' + 2s^2 \cdot \varepsilon^{\mathsf{REnc}}(k, s) \\
&\leq 2s \cdot \eta(k) + 2s((4s^2 + 2)2^{-\kappa/2} + s \cdot \varepsilon^{\mathsf{REnc}}(k, s)) + 2s \cdot \varepsilon^{\mathsf{REnc}}(k, s) \\
&\leq 2s \cdot \eta(k) + 4s^2 \cdot \varepsilon^{\mathsf{REnc}}(k, s) + (8s^3 + 4s)2^{-\kappa(k)/2}
\end{aligned}$$

■

5.3 Chosen-ciphertext secure construction in the standard model

In this section, we present a scheme which is secure against chosen-ciphertext attack. It is essentially the scheme of [RSV13a] and is detailed in Algorithm 5.15.

Theorem 5.16. *Fix any $\kappa(k) \geq 2a + 2\log t_2 + \omega(\log k)$. The encryption system in Algorithm 5.15 is* Π-$\mathsf{ROR\text{-}SCDCA}$-*blockwise-secure for* $\Pi = (\mathrm{H}_\infty(\mathbf{pk}) - b, \kappa, \kappa)$.

The security reduction is very close to the one given in [RSV13a]. We consider a sequence of games $\mathcal{G}^{(i)}$ obtained from the security game $\mathsf{ROR\text{-}SCDCA}$ by modifying the challenge ciphertext generation for $i \in \{0, \dots, T\}$ and modify these games into $\mathcal{G}_0^{(i)}, \mathcal{G}_1^{(i)}, \dots, \mathcal{G}_6^{(i)}$ in a way very similar to [RSV13a].

We note that our construction works for slightly superpolynomial T because t_1, t_2 depend on T logarithmically, and therefore our scheme satisfies security against all polynomial T (with no *a priori* bound on

Fix any polynomials $a(k), b(k), n(k), \ell(k)$ such that $\ell(k) = \omega(a(k))$ and $a(k) = (n(k) + \ell(k))^{\Omega(1)}$. Fix a family of $(n+\ell, a)$-lossy trapdoor functions. Fix superpolynomial $T = k^{\omega(1)}$ and set $\nu = \sqrt{a}$, $t_1 = b+n+\nu+2\log T+\omega(\log k)$, $\delta_1 = 2^{-(n+\ell)t_1}$, $t_2 = b+n+2a+2\log T+\omega(\log k)$, $\delta_2 = 2^{-(n+\ell)t_2}$. We define the following public-key encryption:

Key generation : $\mathsf{Gen}(1^k)$ samples $h : \{0,1\}^{n+\ell} \to \{0,1\}^{\nu}$ from an admissible hash function family $\mathcal{H}$. $\mathsf{Gen}(1^k)$ runs the injective generation function for a $(n+\ell, a)$-lossy trapdoor function family to generate (f, f^{-1}). $\mathsf{Gen}(1^k)$ runs the generation function for a $\mathcal{R}^{BM}$-$(n+\ell, a)$-lossy trapdoor function family with a uniformly chosen initialization parameter K to generate (g, g^{-1}). $\mathsf{Gen}(1^k)$ also samples $\pi_1 \leftarrow_{\mathrm{R}} \boldsymbol{\pi}$ from a t_1-wise δ_1-dependent permutation over $\{0,1\}^{n+\ell}$. and $\pi_2 \leftarrow_{\mathrm{R}} \boldsymbol{\pi}$ from a t_2-wise δ_2-dependent permutation over $\{0,1\}^{n+\ell}$. The public key is $\mathrm{pk} = (h, f, g, \pi_1, \pi_2)$ and the private key is $\mathrm{sk} = (f^{-1}, g^{-1})$.

Encryption : $\mathsf{Enc}_{\mathrm{pk}}(m; r) = (h(\pi_1(m\|r)), f(\pi_2(m\|r)), g(h(\pi_1(m\|r)), \pi_2(m\|r)))$ where $r \in \{0,1\}^{\ell(k)}$.

Decryption : $\mathsf{Dec}_{\mathrm{sk}}(c_1, c_2, c_3) = \pi_2^{-1}(f^{-1}(c_2))_{[n]}$ if $c_1 = h(\pi_1(\pi_2^{-1}(f^{-1}(c))))$ and $c_3 = g(c_1, f^{-1}(c_2))$. $\mathsf{Dec}_{\mathrm{sk}}(c_1, c_2, c_3) = \perp$ otherwise.

Algorithm 5.15. ROR-SCDCA encryption in standard model

the polynomial). This is an improvement over [RSV13a], which obtain key size that is *linear* in T, and therefore one can only achieve security against T that is a polynomial fixed ahead of time.[1] We are able to improve this dependence by observing that at one step of the proof of security, [RSV13a] uses Lemma 4.5 of [RSV13a] (restated in Lemma 2.6) but pays heavily due to the fact that this lemma requires the high conditional entropy condition to hold always. We prove the following average-case version of Lemma 4.5 of [RSV13a], and note that this average-case version suffices for the proof of security, while allowing us to save in the key size.

Theorem 5.17. *For any $\varepsilon, \gamma, \eta \in (0,1)$, for any $(n+\ell, a)$-lossy function $f : \{0,1\}^{n+\ell} \to \{0,1\}^{n'}$, and for $\boldsymbol{\pi}$ a t-wise δ-dependent permutation over $\{0,1\}^{n+\ell}$ where $t = b + a + \log(1/\varepsilon) + \log(1/\eta) + 1$ and $\delta \leq 2^{-(n+\ell)t}$, the following holds:*
Let $\mathcal{R} = \{(\mathbf{x}, \mathbf{y})\}$ be a family of sources of size $|\mathcal{R}| \leq 2^b$ satisfying that for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{R}$, it holds that $\Pr_{x \xleftarrow{R} \mathbf{x}}[\mathrm{H}_\infty(\mathbf{y} \mid \mathbf{x} = x) < \kappa] < \gamma$ where $\kappa = a + 2\log(1/\varepsilon) + 2\log t + \Theta(1)$.
Then with probability at least $1 - \eta$ over the choice of $\pi \in \boldsymbol{\pi}$, it holds for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{R}$ that:

$$\Delta\left((\mathbf{x}, f(\pi(\mathbf{y}))), (\mathbf{x}, f(\pi(\mathbf{u}_{n+\ell})))\right) \leq \varepsilon + \gamma$$

where $\mathbf{u}_{n+\ell}$ is an independant uniform string over $\{0,1\}^{n+\ell}$.

Proof. Fix any particular $(\mathbf{x}, \mathbf{y}) \in \mathcal{R}$. Define the event $\mathsf{Low}(x)$ to be those x where $\mathrm{H}_\infty(\mathbf{y} \mid \mathbf{x} = x) < \kappa$ and observe that by hypothesis, it holds that $\Pr[\mathsf{Low}(\mathbf{x})] < \gamma$.

[1] Earlier versions of [RSV13a] also claimed a dependence of $\log T$ but this was in fact a typographical error, and the currently available ePrint archive version contains the rectified dependence, which is linear in T.

Observe that, for fixed π

$$\begin{aligned}&\Delta((\mathbf{x}, f(\pi(\mathbf{y}))), (\mathbf{x}, f(\mathbf{u})))\\ &\quad \le \Pr[\mathsf{Low}(\mathbf{x})] \;+\; \Delta(\;(\mathbf{x}, f(\pi(\mathbf{y})) \mid \neg\mathsf{Low}(\mathbf{x})),\; (\mathbf{x}, f(\mathbf{u}) \mid \neg\mathsf{Low}(\mathbf{x}))\;)\\ &\quad < \gamma \;+\; \Delta(\;(\mathbf{x}, f(\pi(\mathbf{y})) \mid \neg\mathsf{Low}(\mathbf{x})),\; (\mathbf{x}, f(\mathbf{u}) \mid \neg\mathsf{Low}(x))\;)\end{aligned}$$

By Lemma 2.6, it holds that with probability $\ge 1 - 2^{a-t+1}/\varepsilon$ over the choice of π that

$$\Delta((\mathbf{x}, f(\pi(\mathbf{y})) \mid \neg\mathsf{Low}(\mathbf{x})), (\mathbf{x}, f(\mathbf{u}) \mid \neg\mathsf{Low}(x))) \le \varepsilon$$

Plugging in our parameters gives

$$\Pr_{\pi \xleftarrow{R} \boldsymbol{\pi}}[\Delta((\mathbf{x}, f(\pi(\mathbf{y}))), (\mathbf{x}, f(\mathbf{u}))) > \gamma + \varepsilon] \le 2^{-b}\eta$$

Taking a union bound over all 2^b distributions implies the theorem. ■

We also note that Theorem 5.17 can also be plugged back into the proof of [RSV13a, Theorem 7.1] to obtain a key size that is linear in $\log T$ for their scheme.

5.3.1 Security proof for Algorithm 5.15

Proof of Theorem 5.16. We strengthen the security game in the same way as in the proof of Theorem 5.5: instead of the original RoR oracle, we use the analogue RoR′ oracle from Theorem 5.5: in real mode it behaves just as RoR, while in random mode it encrypts random messages using the same lossy functions f, g, h but with *independently* chosen permutations $\boldsymbol{\pi}'_1, \boldsymbol{\pi}'_2$. Call the resulting game ROR-SCDCA′. As in Theorem 5.5 security in ROR-SCDCA′ implies security in ROR-SCDCA.

Let $\mathcal{A} = (\mathsf{Rev}, \mathsf{Query}, \mathsf{Mesg}, \mathsf{Rand})$ be a ROR-SCDCA′-adversary. We assume that $\mathcal{A}$ always makes exactly q queries (for some polynomial $q(k)$) decryption queries and that Mesg and Rand always output vectors (m^*, r^*) of length T (for some polynomial $T(k)$) in the security game ROR-SCDCA′. We assume (for contradiction) that $\mathcal{A}$ has advantage ϵ for some non-negligible function $\epsilon(k)$.

For any polynomial q, security for adversaries making q queries to the RoR′ oracle in the ROR-SCDCA′ game is equivalent to security for adversaries making a single query to the RoR′ oracle, and so in the rest of the proof we assume that the adversary makes only a single RoR′ query (as in the proof of Theorem 5.5).

We denote by $c^{(1)}, \dots c^{(q)}$ the random variable corresponding to these decryption queries and by $\mathbf{c}^* = (c_1^*, \dots, c_T^*)$ the vector of random variables corresponding to the challenge ciphertext returned by the RoR′ oracle.

For $i \in \{0, \dots, T\}$, we consider the following sequence of games $\mathcal{G}^{(i)}$ obtained from the security game ROR-SCDCA′ by modifying the challenge ciphertext generation. In the real mode of the ROR-SCDCA′ experiment, the oracle RoR′ is given the tuple (m^*, r^*) generated by the algorithms Mesg and Rand and outputs the ciphertext $c^* = (c_1^*, \dots, c_T^*) = (\mathsf{Enc}_{\mathrm{pk}}(m_1^*, r_1^*), \dots, \mathsf{Enc}_{\mathrm{pk}}(m_T^*, r_T^*))$. In $\mathcal{G}^{(i)}$, the RoR′ oracle samples uniform $u_{T-i+1}, \dots, u_T \xleftarrow{R} \{0,1\}^n$ and $s_{T-i+1}, \dots, s_T \xleftarrow{R} \{0,1\}^\ell$ and uniform independent permutations π'_1, π'_2 and outputs the challenge ciphertext

$$c^* = (c_1^*, \dots, c_T^*) = (\mathsf{Enc}_{\mathrm{pk}}(m_1^*, r_1^*), \dots, \mathsf{Enc}_{\mathrm{pk}}(m_{T-i}^*, r_{T-i}^*), \mathsf{Enc}_{\mathrm{pk}'}(u_{T-i+1}^*, s_{T-i+1}^*), \dots, \mathsf{Enc}_{\mathrm{pk}'}(u_T^*, s_T^*)).$$

where $\mathrm{pk}' = (f, g, h, \pi_1', \pi_2')$. In particular, we have $\mathcal{G}^{(0)} = (\mathsf{ROR\text{-}SCDCA}')^{\mathsf{real}}$ and $\mathcal{G}^{(T)} = (\mathsf{ROR\text{-}SCDCA}')^{\mathsf{random}}$ and we will prove that

$$|\Pr[\mathcal{G}^{(i)}(k) = 1] - \Pr[\mathcal{G}^{(i+1)}(k) = 1]| \tag{5.10}$$

is smaller than $\epsilon(k)/T(k)$ for all $i \in \{0, \ldots, T-1\}$ and for a security parameter k large enough (and therefore reach the contradiction).

In the following, we will modify the games $\mathcal{G}^{(i)}(k)$ and $\mathcal{G}^{(i+1)}(k)$ for $i \in \{0, \ldots, T-1\}$. We fix the value of $i \in \{0, \ldots, T-1\}$ and let $j \in \{i, i+1\}$.

Game $\mathcal{G}_0^{(j)}(k)$ The games $\mathcal{G}_0^{(j)}(k)$ for $j \in \{i, i+1\}$ are identical to $\mathcal{G}^{(j)}(k)$ except that the initialization parameter K is sampled from $\mathcal{K}_{u,\nu}$ as for the polynomial u as promised to exist by Definition 2.4. Since no adversary can distinguish between different initialization parameters (Definition 2.3), it holds that

$$|\Pr[\mathcal{G}^{(j)}(k) = 1] - \Pr[\mathcal{G}_0^{(j)}(k) = 1]| \leq \mathsf{negl}(k)$$

Game $\mathcal{G}_1^{(j)}(k)$ The games $\mathcal{G}_1^{(j)}(k)$ for $j \in \{i, i+1\}$ are identical to $\mathcal{G}_0^{(j)}(k)$ except that they output 0 whenever the tuple corresponding to (m_{T-i}, r_{T-i}) and the message/randomness pairs corresponding to the decryption queries forms a bad sequence of inputs for the admissible hash function h.

Let $x^* = \pi_1(m^*_{T-i}, r^*_{T-i})$ and $x_t = \pi_1(\mathsf{Dec}_{\mathrm{sk}}(c^t))$ for $t \in \{1, \ldots, q\}$. The games $\mathcal{G}_1^{(j)}(k)$ output whatever $\mathcal{G}_0^{(j)}(k)$ does when $(x^*, x_1, \ldots, x_q) \notin \mathsf{Unlikely}_h$ and aborts and outputs a random bit when $(x^*, x_1, \ldots, x_q) \in \mathsf{Unlikely}_h$. Since the probability that $(x^*, x_1, \ldots, x_q) \in \mathsf{Unlikely}_h$ is negligible, we have readily

$$|\Pr[\mathcal{G}_0^{(j)}(k) = 1] - \Pr[\mathcal{G}_1^{(j)}(k) = 1]| \leq \mathsf{negl}(k).$$

Game $\mathcal{G}_2^{(j)}(k)$ The games $\mathcal{G}_2^{(j)}(k)$ for $j \in \{i, i+1\}$ are obtained from $\mathcal{G}_1^{(j)}(k)$ by outputting the output of $\mathcal{G}_1^{(j)}(k)$ with probability $\Gamma(k)^{-1}$ and aborting and outputting a random bit with probability $1 - \Gamma(k)^{-1}$ (where $\Gamma(k)$ is the probability from the admissible hash function definition). We have

$$\Pr[\mathcal{G}_2^{(j)}(k) = 1] = \frac{1}{\Gamma}\Pr[\mathcal{G}_1^{(j)}(k) = 1] + \left(1 - \frac{1}{\Gamma}\right)\frac{1}{2}$$

for $j \in \{i, i+1\}$ and therefore

$$|\Pr[\mathcal{G}_2^{(i+1)}(k) = 1] - \Pr[\mathcal{G}_2^{(i)}(k) = 1]| = \frac{1}{\Gamma}|\Pr[\mathcal{G}_1^{(i+1)}(k) = 1] - \Pr[\mathcal{G}_1^{(i)}(k) = 1]|.$$

Game $\mathcal{G}_3^{(j)}(k)$ This game executes $\mathcal{G}_1^{(j)}(k)$ and then does the following if it did not abort. It samples an independent initialization key $K' \overset{R}{\leftarrow} \mathcal{K}_{u,\nu}$ for the $\mathcal{R}$-lossy trapdoor function family (in addition to the actual key K used in the key generation). We denote $\mathsf{Partition}_{K,h}^{(j)}$ the event in which $P_{K'}(h(x^*)) = \mathtt{Lossy}$ and $P_{K'}(h(x_t)) = \mathtt{Inj}$ for $t \in \{1, \ldots, q\}$. By definition of the ROR-SCDCA$'$ security games, $x^* \notin \{x_1, \ldots, x_q\}$ and we know that $\Pr[\mathsf{Partition}_{K',h}^{(j)}] \geq \Gamma^{-1}$ since the game did not yet abort. Next we approximate the actual probability $\Pr[\mathsf{Partition}_{K',h}^{(j)}]$ by sampling $\lceil kS \cdot \Gamma \rceil$ keys for the $\mathcal{R}$-lossy trapdoor function family (for some polynomial $S(k)$ that will be determined at the end of the proof). Using Hoeffding's inequality, obtain an approximation $\tilde{p}^{(j)}$ of $\Pr[\mathsf{Partition}_{K',h}^{(j)}]$ such that

$$\Pr\left[|\Pr[\mathsf{Partition}_{K',h}^{(j)}] - \tilde{p}^{(j)}| \geq \frac{1}{S \cdot \Gamma}\right] \leq \frac{1}{2^k}.$$

Finally, if $\mathsf{Partition}^{(j)}_{K',h}$ occurs then the game outputs the same as the output of $\mathcal{G}^{(j)}_1(k)$ with probability $(\Gamma\tilde{p}^{(j)})^{-1})$. In all other cases it aborts and outputs a random bit. We have

$$|\Pr[\mathcal{G}^{(j)}_3(k)=1]-\Pr[\mathcal{G}^{(j)}_2(k)=1]| \leq \frac{1}{\Gamma\cdot S}+\frac{1}{2^k}$$

Game $\mathcal{G}^{(j)}_4(k)$ The games $\mathcal{G}^{(j)}_4(k)$ for $j\in\{i,i+1\}$ are identical to $\mathcal{G}^{(j)}_3(k)$ except that we replace the events $\mathsf{Partition}^{(j)}_{K',h}$ by the event $\mathsf{Partition}^{(j)}_{K,h}$. An adversary able to distinguish these games can be used to distinguish initialization keys for the $\mathcal{R}$-lossy trapdoor function family , so we have

$$|\Pr[\mathcal{G}^{(j)}_4(k)=1]-\Pr[\mathcal{G}^{(j)}_3(k)=1]| \leq \mathsf{negl}(k).$$

Game $\mathcal{G}^{(j)}_5(k)$ In the games $\mathcal{G}^{(j)}_4(k)$ for $j\in\{i,i+1\}$, the decryption queries are answered using the trapdoor g^{-1} for the $\mathcal{R}$-lossy trapdoor function g instead of the trapdoor f^{-1} for the lossy trapdoor function f. We have

$$\Pr[\mathcal{G}^{(j)}_5(k)=1]=\Pr[\mathcal{G}^{(j)}_4(k)=1]$$

for $j\in\{i,i+1\}$.

Game $\mathcal{G}^{(j)}_6(k)$ In these final games, the key generation of the encryption scheme is modified by sampling a lossy trapdoor function f instead of an injective one. By the indistinguishability of the two setting (and since the trapdoor f^{-1} is no longer used in games $\mathcal{G}^{(j)}_4(k)$ for $j\in\{i,i+1\}$), we have

$$|\Pr[\mathcal{G}^{(j)}_6(k)=1]-\Pr[\mathcal{G}^{(j)}_5(k)=1]| \leq \mathsf{negl}(k).$$

We have

$$\begin{aligned}
|\Pr[\mathcal{G}^{(i)}(k)=1]-\Pr[\mathcal{G}^{(i+1)}(k)=1]| &\leq |\Pr[\mathcal{G}^{(i)}_0(k)=1]-\Pr[\mathcal{G}^{(i+1)}_0(k)=1]|+\mathsf{negl}(k)\\
&\leq |\Pr[\mathcal{G}^{(i)}_1(k)=1]-\Pr[\mathcal{G}^{(i+1)}_1(k)=1]|+\mathsf{negl}(k)\\
&\leq \Gamma|\Pr[\mathcal{G}^{(i)}_2(k)=1]-\Pr[\mathcal{G}^{(i+1)}_2(k)=1]|+\mathsf{negl}(k)\\
&\leq 2\left(\tfrac{1}{S}+\tfrac{\Gamma}{2^k}\right)+\Gamma|\Pr[\mathcal{G}^{(i)}_3(k)=1]-\Pr[\mathcal{G}^{(i+1)}_3(k)=1]|+\mathsf{negl}(k)\\
&\leq 2\left(\tfrac{1}{S}+\tfrac{\Gamma}{2^k}\right)+\Gamma|\Pr[\mathcal{G}^{(i)}_4(k)=1]-\Pr[\mathcal{G}^{(i+1)}_4(k)=1]|+\mathsf{negl}(k)\\
&\leq 2\left(\tfrac{1}{S}+\tfrac{\Gamma}{2^k}\right)+\Gamma|\Pr[\mathcal{G}^{(i)}_5(k)=1]-\Pr[\mathcal{G}^{(i+1)}_5(k)=1]|+\mathsf{negl}(k)\\
&\leq 2\left(\tfrac{1}{S}+\tfrac{\Gamma}{2^k}\right)+\Gamma|\Pr[\mathcal{G}^{(i)}_6(k)=1]-\Pr[\mathcal{G}^{(i+1)}_6(k)=1]|+\mathsf{negl}(k)
\end{aligned}$$

To conclude the proof, we need to prove that

$$|\Pr[\mathcal{G}^{(i)}_6(k)=1]-\Pr[\mathcal{G}^{(i+1)}_6(k)=1]| \tag{5.11}$$

is negligible in the security parameter k. Indeed, if this the case, it suffices to pick $S(k)$ a polynomial large enough such that

$$2\left(\frac{1}{S}+\frac{\Gamma}{2^k}\right)+\mathsf{negl}(k)\leq\frac{\epsilon(k)}{T(k)}$$

for k sufficiently large (which is possible since ϵ is non-negligible).

To prove that (5.11) is negligible, we analyze the games $\mathcal{G}^{(i)}_6(k)$ and $\mathcal{G}^{(i+1)}_6(k)$. From the simulation, the output of these two random experiments is a uniform bit if $(x^*,x_1,\ldots,x_q)\in\mathsf{Unlikely}_h$, or if $(x^*,x_1,\ldots,x_q)\notin\mathsf{Unlikely}_h$ and the event $\mathsf{Partition}^{(j)}_{K,h}$ does not occur, or if the experiment aborts. Therefore, we consider the event Good defined as the conjunction of these events:

1. $(x^*, x_1, \ldots, x_q) \notin \mathsf{Unlikely}_h$;

2. $\mathsf{Partition}^{(j)}_{K,h}$;

3. the experiment does not abort.

The challenge ciphertext $c^* = (c^*_1, \ldots, c^*_T)$ in the two experiments $\mathcal{G}^{(i)}_6(k)$ and $\mathcal{G}^{(i+1)}_6(k)$ differs only on the $(T-i)$-th position. In $\mathcal{G}^{(i)}_6(k)$, c^*_{T-i} is the encryption of m^*_{T-i} using random coins r^*_{T-i} whereas in $\mathcal{G}^{(i)}_6(k)$ it is the encryption of u_{T-i} using random coins s_{T-i} and independent permutations π'_1, π'_2. More precisely, c^*_{T-i} is equal to

$$\begin{cases} (c_h, c_f, c_g) & := & (h(\pi_1(m_{T-i}\|r_{T-i})), f(\pi_2(m_{T-i}\|r_{T-i})), g(h(\pi_1(m_{T-i}\|r_{T-i})), \pi_2(m_{T-i}\|r_{T-i}))), \text{if } j = i \\ (u_h, u_f, u_g) & := & (h(\pi'_1(u_{T-i}\|s_{T-i})), f(\pi'_2(u_{T-i}\|s_{T-i})), g(h(\pi'_1(u_{T-i}\|s_{T-i})), \pi'_2(u_{T-i}\|s_{T-i}))), \text{if } j = i+1 \end{cases}$$

We denote $\mathbf{c}_h, \mathbf{c}_f, \mathbf{c}_g$ the random variables corresponding to c_h, c_f, c_g in $\mathcal{G}^{(i)}_6(k)$ and $\mathbf{u}_h, \mathbf{u}_f, \mathbf{u}_g$ the random variables corresponding to u_h, u_f, u_g in $\mathcal{G}^{(i)}_6(k)$. We need to prove that the two distributions $(\mathbf{c}_h, \mathbf{c}_f, \mathbf{c}_g)$ and $(\mathbf{u}_h, \mathbf{u}_f, \mathbf{u}_g)$ are statistically close. We distinguish the following two cases.

If $p \leq b$: we assume *w.l.o.g.* that Query_k is deterministic, since it can always push any randomness it uses into its query Mesg. In this case the input to Query_k can be one of at most 2^b values, and therefore the output distribution $(\boldsymbol{\xi}, \mathbf{m}, \mathbf{r})$ of the interaction of $\mathsf{Mesg}, \mathsf{Rand}$ can be one of at most 2^b possible distributions. Furthermore, by the third entropy condition of Definition 3.2, $(\mathbf{m}, \mathbf{r})$ form a κ-source, even when conditioned on the values of ρ, ξ.

Note that since $\mathbf{mr}$ is a κ-source with overwhelming probability, it holds that with overwhelming probability over the choice of $(mr)_{[T-i-1]}$, it holds that

$$\mathrm{H}_\infty\left((\mathbf{m}^*_{T-i}, \mathbf{r}^*_{T-i}) | (\mathbf{mr})_{[T-i-1]} = (mr)_{[T-i-1]}, \boldsymbol{\xi} = \xi, \boldsymbol{\rho} = \rho\right) \geq \kappa$$

Therefore, Theorem 5.17 implies that, even given $\rho, \xi, (mr)_{[T-i-1]}$, it holds that for all 2^b possible distributions of $\mathbf{c}_h = h(\boldsymbol{\pi}_1(\mathbf{m}_{T-i}\|\mathbf{r}_{T-i}))$, it is indistinguishable from $\mathbf{u}_h = h(\boldsymbol{\pi}'_1(\mathbf{u}\|\mathbf{s}))$ for uniform random $\mathbf{u}, \mathbf{s}$ of appropriate length and any value of $\boldsymbol{\pi}'_1$ since it is a permutation.

Furthermore, since the output of h is of length a, we may apply Lemma 2.1 and deduce that with overwhelming probability over the choice of $\mathbf{c}_h$ that:

$$\mathrm{H}_\infty\left((\mathbf{m}^*_{T-i}, \mathbf{r}^*_{T-i}) | \mathbf{c_h} = c_h, (\mathbf{mr})_{[T-i-1]} = (mr)_{[T-i-1]}, \boldsymbol{\xi} = \xi, \boldsymbol{\rho} = \rho\right) \geq \kappa - \nu - \omega(\log k)$$

Therefore, Theorem 5.17 implies that, even given $\rho, \xi, (mr)_{[T-i-1]}, c_h$, it holds that for all possible 2^b choices of

$$(\mathbf{c}_f, \mathbf{c}_g) = (f(\boldsymbol{\pi}_2(\mathbf{m}_{T-i}\mathbf{r}_{T-i})), g(c_h, \boldsymbol{\pi}_2(\mathbf{m}_{T-i}\mathbf{r}_{T-i})))$$

that it is indistinguishable from

$$(\mathbf{u}_f, \mathbf{u}_g) = (f(\boldsymbol{\pi}'_2(\mathbf{u}\|\mathbf{s})), g(c_h, \boldsymbol{\pi}'_2(\mathbf{u}\|\mathbf{s})))$$

where $\mathbf{u}, \mathbf{s}$ are independent randomness of appropriate length. We may apply Theorem 5.17 because $x \mapsto (f(x), g(c_h, x))$ is a shrinking function mapping $n + \ell$ bits to $2a$ bits.

If $p > b$: in this case, the first entropy condition of Definition 3.2 implies that $c \leq b$. Furthermore, the second entropy condition implies that $\mathbf{r}_\sigma$ is a κ-block-wise-source with overwhelming probability over σ.

$\mathcal{G}_6^{(i+1)}(k)$ is like $\mathcal{G}_6^{(i)}(k)$ except the the $(T-i)$-th ciphertext output by the RoR' oracle is uniform.

The view of the adversary in $\mathcal{G}_6^{(i)}(k)$ is the tuple

$$\mathbf{X}_i = (\mathbf{h}, \mathbf{f}, \mathbf{g}, \boldsymbol{\xi}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \mathbf{c}_1, \ldots, \mathbf{c}_{T-i-1}, (\mathbf{c}_h, \mathbf{c}_f, \mathbf{c}_g), \mathbf{c}_{T-i+1}, \ldots, \mathbf{c}_T)$$

while its view in $\mathcal{G}_6^{(i+1)}(k)$ is the tuple

$$\mathbf{X}_{i+1} = (\mathbf{h}, \mathbf{f}, \mathbf{g}, \boldsymbol{\xi}, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \mathbf{c}_1, \ldots, \mathbf{c}_{T-i-1}, (\mathbf{u}_h, \mathbf{u}_f, \mathbf{u}_g), \mathbf{c}_{T-i+1}, \ldots, \mathbf{c}_T)$$

where

$$\mathbf{c}_j = (\mathbf{h}(\boldsymbol{\pi}_1(\mathbf{m}_i \| \mathbf{r}_i), \mathbf{f}(\boldsymbol{\pi}_2(\mathbf{m}_i \| \mathbf{r}_i), \mathbf{g}(\mathbf{h}(\boldsymbol{\pi}_1(\mathbf{m}_i \| \mathbf{r}_i)), \boldsymbol{\pi}_2(\mathbf{m}_i \| \mathbf{r}_i)))$$

for $j \in \{1, \ldots, T-i-1\}$ and

$$\mathbf{c}_j = (\mathbf{h}(\boldsymbol{\pi}_1(\mathbf{u}_i \| \mathbf{s}_i), \mathbf{f}(\boldsymbol{\pi}_2(\mathbf{u}_i \| \mathbf{s}_i), \mathbf{g}(\mathbf{h}(\boldsymbol{\pi}_1(\mathbf{u}_i \| \mathbf{s}_i)), \boldsymbol{\pi}_2(\mathbf{u}_i \| \mathbf{s}_i)))$$

for $j \in \{T-i+1, \ldots, T\}$. The first $(T-i+1)$ components and the last i components in both distributions are identical.

Suppose now we reveal $\mathbf{m}$ and $\boldsymbol{\sigma}$ to the adversary for free (this can only increase statistical distance, so bounding the distance with this additional revealed information suffices). Since there are at most 2^b possible transcripts σ, $\mathbf{r}_\sigma$ can be one of at most 2^b possible distributions, and due to the second entropy condition of Definition 3.2 and the fact that $\beta = \kappa$ in our attacker profile, it holds that with overwhelming probability over σ, $\mathbf{r}_\sigma$ is a κ-source. We want to bound

$$\Delta((\mathbf{X_i}, \boldsymbol{\sigma}, \mathbf{m}), (\mathbf{X_{i+1}}, \boldsymbol{\sigma}, \mathbf{m}))$$

which, since all coordinates of X_i, X_{i+1} are identical after the $T-i$'th coordinate, is bounded by the distance between $(\mathbf{r}_{[T-i-1]}, \mathbf{c}_h, \mathbf{c}_f, \mathbf{c}_g)$ and $(\mathbf{r}_{[T-i-1]}, \mathbf{u}_f, \mathbf{u}_g, \mathbf{u}_h)$ conditioned on the knowledge of $f, g, h, \xi, \pi_1, \pi_2, \sigma, m$.

It therefore actually suffices then to bound

$$\mathbb{E}_{(f,g,h,\pi_1,\pi_2)}\left[\max_{\sigma,m}\left\{\Delta\left((\mathbf{r}_{[T-i-1]}, \mathbf{c}_h, \mathbf{c}_f, \mathbf{c}_g \mid fgh\pi_1\pi_2\sigma m), (\mathbf{r}_{[T-i-1]}, \mathbf{u}_h, \mathbf{u}_f, \mathbf{u}_g \mid fgh\pi_1\pi_2\sigma m)\right)\right\}\right]$$

(Notice we may ignore ξ for the same reason as in the proof of Theorem 5.5, *i.e.* because the only way Mesg can influence Rand is through their communication σ, and therefore the only dependence of $\mathbf{r}_\sigma$ on ξ is through σ, which is already fixed.)

Observe that, given that σ is fixed, $\mathbf{r}$ is independent of m, and therefore $\mathbf{r}, \mathbf{c}_f, \mathbf{c}_g, \mathbf{c}_h, \mathbf{u}_f, \mathbf{u}_g, \mathbf{u}_h$ are independent of m_j for $j \neq i$. Therefore it suffices to bound

$$\mathbb{E}_{(f,g,h,\pi_1,\pi_2)}\left[\max_{\sigma,m_i}\left\{\Delta\left((\mathbf{r}_{[T-i-1]}, \mathbf{c}_h, \mathbf{c}_f, \mathbf{c}_g \mid fgh\pi_1\pi_2\sigma m_i), (\mathbf{r}_{[T-i-1]}, \mathbf{u}_h, \mathbf{u}_f, \mathbf{u}_g \mid fgh\pi_1\pi_2\sigma m_i)\right)\right\}\right]$$

Therefore there are only 2^{n+b} possible distributions, one for each choice of m_i and σ.

We may conclude the argument by applying Theorem 5.17 as in the previous case of $p \leq b$.

■

6 Discussion and Open Questions

Our definition of Strong CDA security is a *second-degree* assumption in the terminology of Bellare *et al.* [BHK13] due to the fact that the adversary is split into several components and they only have limited communication between them. We show, nevertheless, that our notion is achievable under standard *first-degree* assumptions. This is due to the fact that the kind of separation we have introduced between the different parts of the adversary are *information-theoretic* (by bounding communication or entropy), rather than computational.

One could conceivably relax our model to allow for the kind of computational separation between parts of the adversary, as used in the definition of UCE's [BHK13]. However, it is unlikely that the proofs of security we have presented for our standard-model construction extend to such a relaxed definition, as UCE-type security seem hard to achieve from standard "first-degree" assumptions such as the ones we have used. On the other hand, the random oracle construct we presented would naturally extend; indeed even in its present form the use of the random oracle can be replaced with UCE's to achieve security in the standard model (under the assumption of the existence of UCE's).

Another interesting open question is whether one can improve the parameters of our standard model construction. Our construction requires $\mathrm{H}_\infty(\mathbf{pk}) - b$ entropy in the public key from Rand's point of view (in the first entropy condition of Definition 3.2), which, observing that $\mathrm{H}_\infty(\mathbf{pk}) = \Omega(kb)$ from the construction of t-wise δ-dependent permutations, is a strong requirement. It would be very interesting to construct schemes where this entropy requirement is smaller, perhaps matching the minimal $\omega(\log k)$ requirement we obtain in the random oracle model, or at least achieving k^ε for arbitrarily small constants ε.

7 Acknowledgments

The authors would like to acknowledge helpful comments by Gil Segev and Salil Vadhan. D.V. and D.X. are supported in part by the French ANR JCJC ROMAnTIC project (ANR-12-JS02-0004). D.X. is supported in part by the French ANR Blanc RDAM project (ANR-12-BS02-005).

References

[BB04] Dan Boneh and Xavier Boyen. Secure identity based encryption without random oracles. In Matthew Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 443–459. Springer, August 2004.

[BBDP01] Mihir Bellare, Alexandra Boldyreva, Anand Desai, and David Pointcheval. Key-privacy in public-key encryption. In Colin Boyd, editor, *Advances in Cryptology – ASIACRYPT 2001*, volume 2248 of *Lecture Notes in Computer Science*, pages 566–582. Springer, December 2001.

[BBN$^+$09] Mihir Bellare, Zvika Brakerski, Moni Naor, Thomas Ristenpart, Gil Segev, Hovav Shacham, and Scott Yilek. Hedged public-key encryption: How to protect against bad randomness. In Mitsuru Matsui, editor, *Advances in Cryptology – ASIACRYPT 2009*, volume 5912 of *Lecture Notes in Computer Science*, pages 232–249. Springer, December 2009.

[BBO07] Mihir Bellare, Alexandra Boldyreva, and Adam O'Neill. Deterministic and efficiently searchable encryption. In Alfred Menezes, editor, *Advances in Cryptology – CRYPTO 2007*, volume 4622 of *Lecture Notes in Computer Science*, pages 535–552. Springer, August 2007.

[BCPT13] Eleanor Birrell, Kai-Min Chung, Rafael Pass, and Sidharth Telang. Randomness-dependent message security. In *Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013*, volume 7785 of *Lecture Notes in Computer Science*, pages 700–720. Springer, 2013.

[BFO08] Alexandra Boldyreva, Serge Fehr, and Adam O'Neill. On notions of security for deterministic encryption, and efficient constructions without random oracles. In David Wagner, editor, *Advances in Cryptology – CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 335–359. Springer, August 2008.

[BFOR08] Mihir Bellare, Marc Fischlin, Adam O'Neill, and Thomas Ristenpart. Deterministic encryption: Definitional equivalences and constructions without random oracles. In David Wagner, editor, *Advances in Cryptology – CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 360–378. Springer, August 2008.

[BHK13] Mihir Bellare, Viet Tung Hoang, and Sriram Keelveedhi. Instantiating random oracles via uces. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013*, volume 8043 of *Lecture Notes in Computer Science*, pages 398–415. Springer, 2013.

[BS11] Zvika Brakerski and Gil Segev. Better security for deterministic public-key encryption: The auxiliary-input setting. In Phillip Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *Lecture Notes in Computer Science*, pages 543–560. Springer, August 2011.

[FOR12] Benjamin Fuller, Adam O'Neill, and Leonid Reyzin. A unified approach to deterministic encryption: New constructions and a connection to computational entropy. In Ronald Cramer, editor, *TCC 2012: 9th Theory of Cryptography Conference*, volume 7194 of *Lecture Notes in Computer Science*, pages 582–599. Springer, March 2012.

[Gam85] Taher El Gamal. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, 31(4):469–472, 1985.

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[HO10] Brett Hemenway and Rafail Ostrovsky. Building injective trapdoor functions from oblivious transfer. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:127, 2010.

[KNR09] Eyal Kaplan, Moni Naor, and Omer Reingold. Derandomized constructions of k-wise (almost) independent permutations. *Algorithmica*, 55(1):113–133, 2009.

[MS09] Steven Myers and Abhi Shelat. Bit encryption is complete. In *50th Annual Symposium on Foundations of Computer Science*, pages 607–616. IEEE Computer Society Press, October 2009.

[PW11] Chris Peikert and Brent Waters. Lossy trapdoor functions and their applications. *SIAM J. Comput.*, 40(6):1803–1844, 2011.

[RSV13a] Ananth Raghunathan, Gil Segev, and Salil Vadhan. Deterministic public-key encryption for adaptively chosen plaintext distributions. Cryptology ePrint Archive, Report 2013/125, 2013. Full version of [RSV13b]. Available at http://eprint.iacr.org/.

[RSV13b] Ananth Raghunathan, Gil Segev, and Salil P. Vadhan. Deterministic public-key encryption for adaptively chosen plaintext distributions. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology - EUROCRYPT 2013*, volume 7881 of *Lecture Notes in Computer Science*, pages 93–110. Springer, 2013.

[Vad12] Salil Vadhan. Pseudorandomness (draft survey/monograph), 2012. 220 pages. Available at http://people.seas.harvard.edu/~salil/pseudorandomness/.

[Wee12] Hoeteck Wee. Dual projective hashing and its applications - lossy trapdoor functions and more. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, volume 7237 of *Lecture Notes in Computer Science*, pages 246–262. Springer, April 2012.

[Wic13] Daniel Wichs. Barriers in cryptography with weak, correlated and leaky sources. In Robert D. Kleinberg, editor, *Innovations in Theoretical Computer Science, ITCS '13*, pages 111–126. ACM, 2013.