A New Randomness Extraction Paradigm for Hybrid Encryption

Eike Kiltz[1] Krzysztof Pietrzak[2] Martijn Stam[3] Moti Yung[4]

September 4, 2008

Abstract

We present a new approach to the design of IND-CCA2 secure hybrid encryption schemes in the standard model. Our approach provides an efficient generic transformation from 1-universal to 2-universal hash proof systems. The transformation involves a randomness extractor based on a 4-wise independent hash function as the key derivation function. Our methodology can be instantiated with efficient schemes based on standard intractability assumptions such as DDH, QR and Paillier. Interestingly, our framework also allows to prove IND-CCA2 security of a hybrid version of 1991's Damgård's ElGamal public-key encryption scheme under the DDH assumption.

Keywords: Chosen-ciphertext security, hybrid encryption, randomness extraction, hash proof systems, ElGamal

1 Introduction

Chosen-Ciphertext Security. Indistinguishability against chosen-ciphertext attack (IND-CCA2 security) is by now the accepted standard security definition for public-key encryption schemes. It started with the development of security under lunchtime attacks (also called IND-CCA1) by Naor and Yung [22], who also gave a proof of feasibility using inefficient non-interactive zero-knowledge techniques. This was extended to the more involved systems with IND-CCA2 security in their full generality [24, 9].

Known practical constructions. Efficient designs in the standard model were first presented in the breakthrough works of Cramer and Shoup [3, 4, 5, 26]. At the heart of their design methodology is the notion of *hash proof systems* (HPSs), generalizing the initial system based on the decisional Diffie-Hellman (DDH) problem. Moreover, they are the first to formalize the notion of "Hybrid Encryption," where a public key cryptosystem is used to encapsulate the (session) key of a symmetric cipher which is subsequently used to conceal the data. This is also known as the KEM-DEM approach, after its two constituent parts (the KEM for key encapsulation mechanism, the DEM for data encapsulation mechanism); it is the most efficient way to employ a public key cryptosystem (and encrypting general strings rather than group elements).

Kurosawa and Desmedt [19] later improved upon the original work of Cramer and Shoup with a new paradigm. Whereas Cramer and Shoup [5] require both the KEM and the DEM

[1] CWI Amsterdam, The Netherlands. Email: `kiltz@cwi.nl`. URL: `http://www.cwi.nl/~kiltz`.
[2] CWI Amsterdam, The Netherlands. Email: `pietrzak@cwi.nl`. URL: `www.cwi.nl/~pietrzak`.
[3] EPFL, Switzerland. Email: `martijn.stam@epfl.ch`. URL: `people.epfl.ch/martijn.stam`.
[4] Google Inc. Email: `moti@cs.columbia.edu`. URL: `www1.cs.columbia.edu/~moti/`.

IND-CCA2 secure, Kurosawa and Desmedt show that with a stronger requirement on the DEM (i.e., one-time authenticated encryption), the requirement on the KEM becomes weaker and can be satisfied with any strongly 2-universal hash proof system. (Cramer and Shoup need both a 2-universal and a smooth hash proof system.)

Main Result. The main result of this work is a new paradigm for constructing IND-CCA2 secure hybrid encryption schemes, based on the Kurosawa-Desmedt paradigm. At its core is a surprisingly clean and efficient new method employing *randomness extraction* (as part of the key derivation) to transform a 1-universal hash proof system (that only assures IND-CCA1 security) to a 2-universal hash proof system. From that point on we follow the Kurosawa-Desmedt paradigm: combination with a one-time authenticated encryption scheme (as DEM) will provide IND-CCA2 security of the hybrid encryption scheme.

For the new transformation to work we require a sufficiently compressing 4-wise independent hash function (made part of the public key); we also need a generalization of the leftover hash lemma [15] that may be of independent interest. The efficient transformation enables the design of new and efficient IND-CCA2 secure hybrid encryption schemes based on various hard subset membership problem, such as the DDH assumption, Paillier's DCR assumption, the family of Linear assumptions, and the quadratic residuosity assumption.

A New Proof for Hybrid Damgård's ElGamal. One application of our method is centered around Damgård's public-key scheme [6] (from 1991) which he proved IND-CCA1 secure under the rather strong *knowledge of exponent* assumption.[1] This scheme can be viewed as a "double-base" variant of the original ElGamal encryption scheme [11] and consequently it is often referred to as *Damgård's ElGamal* in the literature. We first view the scheme as a hybrid encryption scheme (as advocated in [26, 5]), applying our methodology of randomness extraction in the KEM's symmetric key derivation before the authenticated encryption (as DEM). The resulting scheme is a hybrid Damgård's ElGamal which is IND-CCA2 secure, under the standard DDH assumption. We furthermore propose a couple of variants of our basic hybrid scheme that offer certain efficiency tradeoffs. Compared to Cramer and Shoup's original scheme [3] and the improved scheme given by Kurosawa-Desmedt [19], our scheme crucially removes the dependence on the hard to construct target collision hash functions (UOWHF), using an easy-to-instantiate 4-wise independent hash function instead.

Related Work. Various previous proofs of variants of Damgård's original scheme have been suggested after Damgård himself proved it IND-CCA1 secure under the strong "knowledge of exponent" assumption (an assumption that has often been criticized in the literature; e.g., it is not efficiently falsifiable according to the classification of Naor [21]). More recent works are by Gjøsteen [14] who showed the scheme IND-CCA1 secure under some *interactive* version of the DDH assumption, where the adversary is given oracle access to some (restricted) DDH oracle. Also, Wu and Stinson [29], and at the same time Lipmaa [20] improve on the above two results. However, their security results are much weaker than ours: they only prove IND-CCA1 security of Damgård's ElGamal, still requiring security assumptions that are either interactive or of "knowledge of exponent" type. Finally, Hieu and Desmedt [7] recently showed a hybrid variant that is IND-CCA2 secure, yet under an even stronger assumption than Damgård's.

We remark that Cramer and Shoup [4] already proposed a generic transformation from 1-universal to 2-universal HPSs. Unfortunately their construction involves a significant overhead: the key of their transformed 2-universal HPS has linearly many keys of the original 1-universal

[1] This assumption basically states that given two group elements (g_1, g_2) with unknown discrete logarithm $\omega = \log_{g_1}(g_2)$, the *only way* to efficiently compute (g_1^x, g_2^x) is to *know* the exponent x.

HPS. We further remark that the notion of randomness extraction has had numerous applications in complexity and cryptography, and in particular in extracting random keys at the final step of key exchange protocols. Indeed, Cramer and Shoup [4] already proposed using a pairwise independent hash function to turn a 1-universal HPS into a 2-universal HPS. Our novel usage is within the context of hybrid encryption as a tool that shifts the integrity checking at decryption time solely to the DEM portion. In stark contrast to the generic transformations by Cramer and Shoup ours is practical. We also remark that several other works also use the general concept of randomness extraction in the setting of public-key cryptography, e.g., [2, 5, 8, 12].

2 Preliminaries

2.1 Notation

If x is a string, then $|x|$ denotes its length, while if S is a set then $|S|$ denotes its size. If $k \in \mathbb{N}$ then 1^k denotes the string of k ones. If S is a set then $s \leftarrow_R S$ denotes the operation of picking an element s of S uniformly at random. We write $\mathsf{A}(x, y, \ldots)$ to indicate that A is an algorithm with inputs $x, y, \ldots$ and by $z \leftarrow_R \mathsf{A}(x, y, \ldots)$ we denote the operation of running A with inputs $(x, y, \ldots)$ and letting z be the output. We write $\lg x$ for logarithms over the reals with base 2. The *statistical distance* between two random variables X and Y having a common domain $\mathcal{X}$ is $\mathsf{SD}(X, Y) = \frac{1}{2}\sum_{x \in \mathcal{X}} |\Pr[X = x] - \Pr[Y = x]|$. The *min-entropy* of a random variable A is defined as $H_\infty(A) = -\lg(\max_{a \in D} \Pr[A = a])$.

2.2 Public-Key Encryption

A *public key encryption* scheme $\mathsf{PKE} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ with message space $\mathcal{M}(k)$ consists of three polynomial time algorithms (PTAs), of which the first two, Kg and Enc, are probabilistic and the last one, Dec, is deterministic. Public/secret keys for security parameter $k \in \mathbb{N}$ are generated using $(pk, sk) \leftarrow_R \mathsf{Kg}(1^k)$. Given such a key pair, a message $m \in \mathcal{M}(k)$ is encrypted by $C \leftarrow_R \mathsf{Enc}(pk, m)$; a ciphertext is decrypted by $m \leftarrow_R \mathsf{Dec}(sk, C)$, where possibly Dec outputs $\perp$ to denote an invalid ciphertext. For consistency, we require that for all $k \in \mathbb{N}$, all messages $m \in \mathcal{M}(k)$, it must hold that $\Pr[\mathsf{Dec}(sk, \mathsf{Enc}(pk, m)) = m] = 1$ where the probability is taken over the above randomized algorithms and $(pk, sk) \leftarrow_R \mathsf{Kg}(1^k)$.

The security we require for PKE is $\mathsf{IND\text{-}CCA2}$ security [24, 10]. We define the advantage of an adversary $\mathsf{A} = (\mathsf{A}_1, \mathsf{A}_2)$ as

$$\mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{PKE},\mathsf{A}}(k) \stackrel{\mathrm{def}}{=} \left| \Pr\left[b = b' \ : \ \begin{array}{l} (pk, sk) \leftarrow_R \mathsf{Kg}(1^k)\ ;\ (m_0, m_1, St) \leftarrow_R \mathsf{A}_1^{\mathsf{Dec}(sk,\cdot)}(pk) \\ b \leftarrow_R \{0,1\}\ ;\ C^* \leftarrow_R \mathsf{Enc}(pk, m_b) \\ b' \leftarrow_R \mathsf{A}_2^{\mathsf{Dec}(sk,\cdot)}(C^*, St) \end{array} \right] - \frac{1}{2} \right| .$$

The adversary A_2 is restricted not to query $\mathsf{Dec}(sk, \cdot)$ with C^*. PKE scheme PKE is said to be indistinguishable against chosen-ciphertext attacks ($\mathsf{IND\text{-}CCA2}$ secure in short) if the advantage function $\mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{PKE},\mathsf{A}}(k)$ is a negligible function in k for all adversaries $\mathsf{A} = (\mathsf{A}_1, \mathsf{A}_2)$ with probabilistic PTA A_1, A_2.

For integers k, t, Q we also define $\mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{PKE},t,Q}(k) = \max_{\mathsf{A}} \mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{PKE},\mathsf{A}}(k)$, where the maximum is over all A that run in time at most t while making at most Q decryption queries.

We also mention the weaker security notion of *indistinguishability against lunch-time attacks* ($\mathsf{IND\text{-}CCA1}$ security), which is defined as $\mathsf{IND\text{-}CCA2}$ security with the restriction that the adversary is not allowed to make decryption queries after having seen the challenge ciphertext.

2.3 Hash Proof Systems

Smooth Projective Hashing. We recall the notion of hash proof systems as introduced by Cramer and Shoup [4]. Let $\mathcal{C}, \mathcal{K}$ be sets and $\mathcal{V} \subset \mathcal{C}$ a language. In the context of public-key encryption (and viewing a hash proof system as a key-encapsulation mechanism (KEM) with special algebraic properties) and may think of $\mathcal{C}$ as the set of all ciphertexts, $\mathcal{V}$ as the set of all consistent ciphertexts, and $\mathcal{K}$ as the set of all symmetric keys. Let $\Lambda_{sk} : \mathcal{C} \to \mathcal{K}$ be a hash function indexed with $sk \in \mathcal{SK}$, where $\mathcal{SK}$ is a set. A hash function Λ_{sk} is *projective* if there exists a projection $\mu : \mathcal{SK} \to \mathcal{PK}$ such that $\mu(sk) \in \mathcal{PK}$ defines the action of Λ_{sk} over the subset $\mathcal{V}$. That is, for every $C \in \mathcal{V}$, the value $K = \Lambda_{sk}(C)$ is uniquely determined by $\mu(sk)$ and C. In contrast, nothing is guaranteed for $C \in \mathcal{C} \setminus \mathcal{V}$, and it may not be possible to compute $\Lambda_{sk}(C)$ from $\mu(sk)$ and C. More precisely, following [17] we define 1- and 2-universal as follows.

1-universal. The projective hash function is ϵ_1-*almost* 1-*universal* if for all $C \in \mathcal{C} \setminus \mathcal{V}$,

$$\mathsf{SD}((pk, \Lambda_{sk}(C)), (pk, K)) \leq \epsilon_1 \tag{1}$$

where in the above $pk = \mu(sk)$ for $sk \leftarrow_R \mathcal{SK}$ and $K \leftarrow_R \mathcal{K}$.

2-universal. The projective hash function is ϵ_2-*almost* 2-*universal* if for all $C, C^* \in \mathcal{C} \setminus \mathcal{V}$ with $C \neq C^*$,

$$\mathsf{SD}((pk, \Lambda_{sk}(C^*), \Lambda_{sk}(C)), (pk, \Lambda_{sk}(C^*), K)) \leq \epsilon_2 \tag{2}$$

where in the above $pk = \mu(sk)$ for $sk \leftarrow_R \mathcal{SK}$ and $K \leftarrow_R \mathcal{K}$.

To a projective hash function we also associate the collision probability, δ, defined as

$$\delta = \max_{C, C^* \in \mathcal{C} \setminus \mathcal{V}, C \neq C^*} (\Pr_{sk}[\Lambda_{sk}(C) = \Lambda_{sk}(C^*)]) \ . \tag{3}$$

Hash Proof System. A hash proof system $\mathsf{HPS} = (\mathsf{Param}, \mathsf{Pub}, \mathsf{Priv})$ consists of three algorithms. The randomized algorithm $\mathsf{Param}(1^k)$ generates parametrized instances of $params = (group, \mathcal{K}, \mathcal{C}, \mathcal{V}, \mathcal{PK}, \mathcal{SK}, \Lambda_{(\cdot)} : \mathcal{C} \to \mathcal{K}, \mu : \mathcal{SK} \to \mathcal{PK})$, where *group* may contain some additional structural parameters. The deterministic public evaluation algorithm Pub inputs the projection key $pk = \mu(sk)$, $C \in \mathcal{V}$ and a witness r of the fact that $C \in \mathcal{V}$ and returns $K = \Lambda_{sk}(C)$. The deterministic private evaluation algorithm Priv inputs $sk \in \mathcal{SK}$ and returns $\Lambda_{sk}(C)$, without knowing a witness. We further assume that μ is efficiently computable and that there are efficient algorithms given for sampling $sk \in \mathcal{SK}$ and sampling $C \in \mathcal{V}$ uniformly together with a witness r.

We say that a hash proof system is 1- (resp. 2-universal) if for all possible outcomes of $\mathsf{Param}(1^k)$ the underlying projective hash function is $\epsilon_1(k)$-almost 1-universal (resp. $\epsilon_2(k)$-almost 2-universal) for negligible $\epsilon_1(k)$ (resp. $\epsilon_2(k)$).

Subset Membership Problem. As computational problem we require that the *subset membership problem* is hard in HPS which means that for random $C_0 \in \mathcal{V}$ and random $C_1 \in \mathcal{C} \setminus \mathcal{V}$ the two elements C_0 and C_1 are computationally indistinguishable. This is captured by defining the advantage function $\mathsf{Adv}^{\mathrm{sm}}_{\mathsf{HPS},\mathsf{A}}(k)$ of an adversary A as

$$\mathsf{Adv}^{\mathrm{sm}}_{\mathsf{HPS},\mathsf{A}}(k) \stackrel{\mathrm{def}}{=} \left| \Pr[\mathsf{A}(\mathcal{C}, \mathcal{V}, C_1) = 1] - \Pr[\mathsf{A}(\mathcal{C}, \mathcal{V}, C_0) = 1] \right|$$

where $\mathcal{C}$ is taken from the output of $\mathsf{Param}(1^k)$, $C_1 \leftarrow_R \mathcal{C}$ and $C_0 \leftarrow_R \mathcal{C} \setminus \mathcal{V}$.

Hash Proof Systems with Trapdoor. Following [19], we also require that the subset membership problem can be efficiently solved with a master trapdoor. More formally, we assume that the hash proof system HPS additionally contains two algorithms Param' and Decide. The alternative parameter generator $\mathsf{Param}'(1^k)$ generates output indistinguishable from the one of $\mathsf{Param}(1^k)$ and additionally returns a trapdoor ω. The subset membership deciding algorithm $\mathsf{Decide}(params, \omega, x)$ returns 1 if $x \in \mathcal{V}$, and 0, otherwise. All known hash proof systems actually have such a trapdoor.

2.4 Symmetric Encryption

A symmetric encryption scheme $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ is specified by its encryption algorithm E (encrypting $m \in \mathcal{M}(k)$ with keys $S \in \mathcal{K}_{\mathsf{SE}}(k)$) and decryption algorithm D (returning $m \in \mathcal{M}(k)$ or $\perp$). Here we restrict ourselves to deterministic algorithms E and D.

The most common notion of security for symmetric encryption is that of (one-time) ciphertext indistinguishability (IND-OT), which requires that all efficient adversaries fail to distinguish between the encryptions of two messages of their choice. Another common security requirement is *ciphertext authenticity.* (One-time) ciphertext integrity (INT-OT) requires that no efficient adversary can produce a new valid ciphertext under some key when given one encryption of a message of his choice under the same key. A symmetric encryption scheme which satisfies *both* requirements simultaneously is called secure in the sense of authenticated encryption (AE-OT secure). Note that AE-OT security is a stronger notion than chosen-ciphertext security. Formal definitions and constructions are provided in Appendix B. There we also recall (following the encrypt-then-mac approach [1, 5]) how to build a symmetric scheme with k-bit keys secure in the sense of AE-OT from the following basic primitives:

- a (computationally secure) one-time symmetric encryption scheme with k-bit keys;
- a (computationally secure) MAC (existentially unforgeable) with k-bit keys;
- and a (computationally secure) key-derivation function.

3 Randomness Extraction

In this section we review a few concepts related to probability distributions and extracting uniform bits from weak random sources. As a technical tool for our new paradigm, we will prove the following generalization of the leftover hash lemma [15]: if $\mathcal{H}$ is 4-wise independent, then $(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X}))$ is close to uniformly random, where $X, \tilde{X}$ can be dependent (but of course we have to require $X \neq \tilde{X}$).

Let $\mathcal{HS}$ be a family of hash functions $\mathcal{H} : \mathcal{X} \to \mathcal{Y}$. With $|\mathcal{HS}|$ we denote the number of functions in this family and when sampling from $\mathcal{HS}$ we assume a uniform distribution. Let $k > 1$ be an integer, the hash-family $\mathcal{HS}$ is k-wise independent if for any sequence of distinct elements $x_1, \ldots, x_k \in \mathcal{X}$ the random variables $\mathcal{H}(x_1), \ldots, \mathcal{H}(x_k)$, where $\mathcal{H} \leftarrow_R \mathcal{HS}$, are uniform random.[2]

Recall that the leftover hash lemma states that for a 2-wise independent hash function $\mathcal{H}$ and a random variable X with min-entropy exceeding the bitlength of $\mathcal{H}$'s range, the random variable $(\mathcal{H}, \mathcal{H}(X))$ is close to uniformly random [15].

Lemma 3.1 Let $X \in \mathcal{X}$ be a random variable where $H_\infty(X) \geq \kappa$. Let $\mathcal{HS}$ be a family of pairwise independent hash functions with domain $\mathcal{X}$ and image $\{0,1\}^\ell$. Then for $\mathcal{H} \leftarrow_R \mathcal{HS}$ and $U_\ell \leftarrow_R \{0,1\}^\ell$

$$\mathsf{SD}((\mathcal{H}, \mathcal{H}(X)), (\mathcal{H}, U_\ell)) \leq 2^{(\ell-\kappa)/2} .$$

[2] A simple construction of a k-wise independent hash function $\mathbb{Z}_p \to \mathbb{Z}_p$ is the following: to sample a function, sample k elements $c_0, \ldots, c_{k-1} \leftarrow_R \mathbb{Z}_p^k$, and define $h_{c_0,\ldots,c_{k-1}}(X) = c_0 + c_1X + c_2X^2 + \ldots + c_{k-1}X^{k-1} \bmod p$.

We will now prove a generalization of the leftover hash lemma that states that even when the hash function is evaluated in two distinct points, the two outputs jointly still look uniformly random. To make this work, we need a 4-wise independent hash function and, as before, sufficient min-entropy in the input distribution. We do note that, unsurprisingly, the loss of entropy compared to Lemma 3.1 is higher, as expressed in the bound on the statistical distance (or alternatively, in the bound on the min-entropy required in the input distribution).

Lemma 3.2 Let $(X, \tilde{X}) \in \mathcal{X} \times \mathcal{X}$ be two random variables (having joint distribution) where $H_\infty(X) \geq \kappa, H_\infty(\tilde{X}) \geq \kappa$ and $\Pr[X = \tilde{X}] = 0$. Let $\mathcal{HS}$ be a family of 4-wise independent hash functions with domain $\mathcal{X}$ and image $\{0,1\}^\ell$. Then for $\mathcal{H} \leftarrow_R \mathcal{HS}$ and $U_{2\ell} \leftarrow_R \{0,1\}^{2\ell}$,

$$\mathsf{SD}((\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})), (\mathcal{H}, U_{2\ell})) \leq 2^{\ell - \kappa/2} .$$

Proof: Let $d = \lg |\mathcal{HS}|$. For a random variable Y and Y' an independent copy of Y, we denote with $Col(Y) = \Pr[Y = Y']$ the collision probability of Y, in particular

$$\begin{aligned}
Col(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) &= \Pr_{\mathcal{H},(X,\tilde{X}),\mathcal{H}',(X',\tilde{X}')}[(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) = (\mathcal{H}', \mathcal{H}'(X'), \mathcal{H}'(\tilde{X}'))] \\
&= \Pr_{\mathcal{H},\mathcal{H}'}[\mathcal{H} = \mathcal{H}'] \cdot \Pr_{\mathcal{H},(X,\tilde{X}),\mathcal{H}',(X',\tilde{X}')}[(\mathcal{H}(X), \mathcal{H}(\tilde{X})) = (\mathcal{H}'(X'), \mathcal{H}'(\tilde{X}')) | \mathcal{H} = \mathcal{H}'] \\
&= \underbrace{\Pr_{\mathcal{H},\mathcal{H}'}[\mathcal{H} = \mathcal{H}']}_{=2^{-d}} \cdot \Pr_{\mathcal{H},(X,\tilde{X}),(X',\tilde{X}')}[(\mathcal{H}(X), \mathcal{H}(\tilde{X})) = (\mathcal{H}(X'), \mathcal{H}(\tilde{X}'))] . \qquad (4)
\end{aligned}$$

We define the event E, which holds if $X, \tilde{X}, X', \tilde{X}'$ are pairwise different.

$$\begin{aligned}
\Pr_{(X,\tilde{X}),(X',\tilde{X}')}[\neg \mathrm{E}] &= \Pr_{(X,\tilde{X}),(X',\tilde{X}')}[X = X' \vee X = \tilde{X}' \vee \tilde{X} = X' \vee \tilde{X} = \tilde{X}'] \\
&\leq 4 \cdot 2^{-\kappa} = 2^{-\kappa+2}
\end{aligned}$$

Where in the first step we used that $X \neq \tilde{X}, X' \neq \tilde{X}'$ by assumption, and in the second step we use the union bound and also our assumption that the min entropy of X and $\tilde{X}$ is at least κ (and thus e.g. $\Pr[X = X'] \leq 2^{-\kappa}$). With this we can write (4) as

$$\begin{aligned}
Col(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) &\leq 2^{-d} \cdot (\Pr[(\mathcal{H}(X), \mathcal{H}(\tilde{X})) = (\mathcal{H}(X'), \mathcal{H}(\tilde{X}')) | \mathrm{E}] + \Pr[\neg \mathrm{E}]) \qquad (5) \\
&\leq 2^{-d}(2^{-2\ell} + 2^{-\kappa+2}) \qquad (6)
\end{aligned}$$

where in the second step we used that $\mathcal{H}$ is 4-wise independent. Let Y be a random variable with support $\mathcal{Y}$ and U be uniform over $\mathcal{Y}$, then

$$\|Y - U\|_2^2 = Col(Y) - |\mathcal{Y}|^{-1}$$

in particular

$$\begin{aligned}
\|(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) - (\mathcal{H}, U_{2\ell})\|_2^2 &= Col(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) - 2^{-d-2\ell} \\
&\leq 2^{-d}(2^{-2\ell} + 2^{-\kappa+2}) - 2^{-d-2\ell} = 2^{-d-\kappa+2}
\end{aligned}$$

Using that $\|Y\|_1 \leq \sqrt{|\mathcal{Y}|}\|Y\|_2$ for any random variable Y with support $\mathcal{Y}$, we obtain

$$\begin{aligned}
\mathsf{SD}((\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})), (\mathcal{H}, U_{2\ell})) &= \frac{1}{2}\|(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) - (\mathcal{H}, U_{2\ell})\|_1 \\
&\leq \frac{1}{2}\sqrt{2^{d+2\ell}}\|(\mathcal{H}, \mathcal{H}(X), \mathcal{H}(\tilde{X})) - (\mathcal{H}, U_{2\ell})\|_2 \\
&\leq \frac{1}{2}\sqrt{2^{d+2\ell}}\sqrt{2^{-d-\kappa+2}} = 2^{\ell-\kappa/2} .
\end{aligned}$$

This concludes the proof. ▮

We note that if $\Pr[X = \tilde{X}] = \delta > 0$, this introduces an additional term of at most δ to the statistical difference above. Moreover, the statement also holds when auxiliary information Z about X and $\tilde{X}$ leaks, as long as $H_\infty(X|Z) \geq \kappa$ and $H_\infty(\tilde{X}|Z) \geq \kappa$ (and $\mathcal{H}$ is independent of $(X, \tilde{X}, Z)$).

4 Hybrid Encryption from Randomness Extraction

In this section we revisit the general construction of hybrid encryption from 2-universal hash proof systems. As our main technical result we show an efficient transformation from a 1-universal to a 2-universal HPS. Combining the latter with an $\mathsf{AE\text{-}OT}$ secure symmetric cipher gives an $\mathsf{IND\text{-}CCA2}$ secure hybrid encryption scheme. This result can be readily applied to all known 1-universal hash proof systems with a hard subset membership problem (e.g., from Paillier, QR [4], DDH, and n-Linear [17, 25]) to obtain a number of new $\mathsf{IND\text{-}CCA2}$ secure hybrid encryption schemes. In Sections 5 and 6 we will work out the consequences for DDH-based schemes.

4.1 Hybrid Encryption from HPS

Recall the notion of a hash proof system from Section 2.3. Kurosawa and Desmedt [19] proposed the following hybrid encryption scheme which improved the schemes from Cramer and Shoup [4].

Let $\mathsf{HPS} = (\mathsf{Param}, \mathsf{Pub}, \mathsf{Priv})$ be a hash proof system and let $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ be an $\mathsf{AE\text{-}OT}$ secure symmetric encryption scheme with key-space $\mathcal{K}$. The system parameters of the scheme consist of $params \leftarrow_R \mathsf{Param}(1^k)$.

$\mathsf{Kg}(k)$**.** Choose random $sk \leftarrow_R \mathcal{SK}$ and define $pk = \mu(sk) \in \mathcal{PK}$. Return (pk, sk).

$\mathsf{Enc}(pk, m)$**.** Pick $C \leftarrow_R \mathcal{V}$ together with its witness r that $C \in \mathcal{V}$. The session key $K = \Lambda_{sk}(C) \in \mathcal{K}$ is computed as $K \leftarrow \mathsf{Pub}(pk, C, r)$. The symmetric ciphertext is $\psi \leftarrow \mathsf{E}_K(m)$. Return the ciphertext (C, ψ).

$\mathsf{Dec}(sk, C)$**.** Reconstruct the key $K = \Lambda_{sk}(C)$ as $K \leftarrow \mathsf{Priv}(sk, C)$ and return $\{m, \perp\} \leftarrow \mathsf{D}_K(\psi)$.

Note that the trapdoor property of the HPS is not used in the actual scheme: it is only needed in the proof. However, as an alternative the trapdoor can be added to the secret key.[3] This allows explicit rejection of invalid ciphertexts during decryption. The security of this explicit-rejection variant is identical to that of the scheme above.

The following was proved in [19, 13, 17].

Theorem 4.1 Assume HPS is ϵ_2-almost 2-universal with hard subset membership problem (with trapdoor), and SE is $\mathsf{AE\text{-}OT}$ secure. Then the encryption scheme is secure in the sense of $\mathsf{IND\text{-}CCA2}$. In particular,

$$\mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{PKE},t,Q}(k) \leq \mathsf{Adv}^{\mathrm{sm}}_{\mathsf{HPS},t}(k) + 2Q \cdot \mathsf{Adv}^{\mathrm{int\text{-}ot}}_{\mathsf{SE},t}(k) + \mathsf{Adv}^{\mathrm{ind\text{-}ot}}_{\mathsf{SE},t}(k) + Q \cdot \epsilon_2 \ .$$

We remark that even though in general the KEM part of the above scheme cannot be proved $\mathsf{IND\text{-}CCA2}$ secure [16], it can be proved "IND-CCCA" secure. The latter notion was defined

[3] Strictly speaking the algorithm to sample elements in $\mathcal{V}$ (with witness) should then be regarded as part of the public key instead of simply a system parameter.

in [17] and proved sufficient to yield IND-CCA2 secure encryption when combined with a AE-OT secure cipher.

There is also an analogue "lite version" for 1-universal HPS, giving IND-CCA1 only (and using a slightly weaker asymmetric primitive). It can be stated as follows.

Theorem 4.2 Assume HPS is 1-universal with hard subset membership problem and SE is WAE-OT secure. Then the encryption scheme is secure in the sense of IND-CCA1.

4.2 A Generic Transformation from 1-Universal to 2-Universal HPS

Our transformation is as follows. Given a projective hash function $\Lambda_{sk} : \mathcal{C} \to \mathcal{K}$ with projection $\mu : \mathcal{SK} \to \mathcal{PK}$ and a family of hash functions $\mathcal{HS}$ with $\mathcal{H} : \mathcal{K} \to \{0,1\}^\ell$. Then we define the hashed variant of it as:

$$\Lambda_{sk}^{\mathcal{HS}} : \mathcal{C} \to \{0,1\}^\ell, \quad \Lambda_{sk}^{\mathcal{HS}}(C) := \mathcal{H}_\tau(\Lambda_{sk}(C)) \ .$$

We also define $\mathcal{PK}^{\mathcal{HS}} = \mathcal{PK} \times \mathcal{HS}$ and $\mathcal{SK}^{\mathcal{HS}} = \mathcal{SK} \times \mathcal{HS}$, such the the hashed projection is given by $\mu^{\mathcal{HS}} : \mathcal{SK}^{\mathcal{HS}} \to \mathcal{PK}^{\mathcal{HS}}$, $\mu^{\mathcal{HS}}(sk, \mathcal{H}) = (pk, \mathcal{H})$. This also induces a transformation from a hash proof system HPS into $\mathsf{HPS}^{\mathcal{HS}}$, where the above transformation is applied to the projective hash function. Note that $\mathcal{C}$ and $\mathcal{V}$ are the same for HPS and $\mathsf{HPS}^{\mathcal{HS}}$ (so that in particular the trapdoor property for the language $\mathcal{V}$ is inherited).

We are now ready to state our main theorem.

Theorem 4.3 Assume HPS is ϵ_1-almost 1-universal with collision probability δ and $\mathcal{HS}$ is a family of 4-wise independent hash functions with $\mathcal{H} : \mathcal{K} \to \{0,1\}^\ell$. Then $\mathsf{HPS}^{\mathcal{HS}}$ is ϵ_2-almost 2-universal for

$$\epsilon_2 = \frac{3}{2} \cdot \frac{2^\ell}{\sqrt{|\mathcal{K}|}} + 3\epsilon_1 + \delta \ .$$

Proof: Let us consider, for all $C, C^* \in \mathcal{C} \setminus \mathcal{V}$ with $C \neq C^*$, the statistical distance relevant for 2-universality for HPS and let Y be the random variable $(pk, \mathcal{H}, U_{2\ell})$ where $pk = \mu(sk)$ for $sk \leftarrow_R \mathcal{SK}$, $\mathcal{H} \leftarrow_R \mathcal{HS}$ and $U_{2\ell} \leftarrow_R \{0,1\}^{2\ell}$. Then we can use the triangle inequality to get

$$\begin{aligned} &\mathsf{SD}((pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), \mathcal{H}(\Lambda_{sk}(C))), (pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), U_\ell)) \\ &\quad \leq \mathsf{SD}((pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), \mathcal{H}(\Lambda_{sk}(C))), Y)) + \mathsf{SD}(Y, (pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), U_\ell)) \end{aligned} \tag{7}$$

where as before $pk = \mu(sk)$ for $sk \leftarrow_R \mathcal{SK}$, $\mathcal{H} \leftarrow_R \mathcal{HS}$ and $U_\ell \leftarrow_R \{0,1\}^\ell$. We can upper bound the second term of (7), using again the triangle inequality in the first step, as

$$\begin{aligned} & \mathsf{SD}(Y, (pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), U_\ell)) \\ \leq\ & \mathsf{SD}(Y, (pk, \mathcal{H}, \mathcal{H}(K), U_\ell)) + \mathsf{SD}((pk, \mathcal{H}, \mathcal{H}(K), U_\ell), (pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*))), U_\ell) \\ \leq\ & \mathsf{SD}(Y, (pk, \mathcal{H}, \mathcal{H}(K), U_\ell)) + \mathsf{SD}((pk, K), (pk, \Lambda_{sk}(C^*))) \\ \leq\ & 2^{\frac{\ell-\kappa}{2}} + \epsilon_1 \ , \end{aligned}$$

where $\kappa = \lg(|\mathcal{K}|)$. In the last step we used the (standard) leftover hash-lemma (Lemma 3.1) and ϵ_1-almost universality of the HPS (cf. (1)) which states that for any $C \in \mathcal{C} \setminus \mathcal{V}$,

$$\mathsf{SD}((pk, K), (pk, \Lambda_{sk}(C))) = \mathsf{SD}(K, \Lambda_{sk}(C) \mid pk) \leq \epsilon_1 \ .$$

By the above, for $C \in \mathcal{C} \setminus \mathcal{V}$ we can define an event E_C, such that $H_\infty(\Lambda_{sk}(C) \mid pk, E_C) = H_\infty(K \mid pk) = \kappa$ where $\Pr[\neg E_C] \leq \epsilon_1$. Further, let E_{Col} denote the event $[\Lambda_{sk}(C) \neq \Lambda_{sk}(C^*)]$, by assumption $\Pr_{sk}[\neg E_{Col}] \leq \delta$.

We now bound the first term of (7) as

$$
\begin{aligned}
& \mathsf{SD}((pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), \mathcal{H}(\Lambda_{sk}(C))), Y) \\
\leq \; & \mathsf{SD}((pk, \mathcal{H}, \mathcal{H}(\Lambda_{sk}(C^*)), \mathcal{H}(\Lambda_{sk}(C))), Y \mid E_C \wedge E_{C^*} \wedge E_{Col}) + \Pr_{sk}[\neg E_C \vee \neg E_{C^*} \vee \neg E_{Col}] \\
\leq \; & 2^{\frac{2\ell - \kappa}{2}} + 2\epsilon_1 + \delta
\end{aligned}
$$

where in the last step we used Lemma 3.2. ▮

4.3 Hybrid Encryption from 1-Universal HPSs

Putting the pieces from the last two section together we get a new IND-CCA2 secure hybrid encryption scheme from any 1-universal hash proof system. Let $\mathsf{HPS} = (\mathsf{Param}, \mathsf{Pub}, \mathsf{Priv})$ be a hash proof system, let $\mathcal{HS}$ be a family of hash functions with $\mathcal{H} : \mathcal{K} \to \{0,1\}^\ell$ and let $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ be an AE-OT secure symmetric encryption scheme with key-space $\{0,1\}^\ell$. The system parameters of the scheme consist of $params \leftarrow_R \mathsf{Param}(1^k)$.

$\mathsf{Kg}(k)$**.** Choose random $sk \leftarrow_R \mathcal{SK}$ and define $pk = \mu(sk) \in \mathcal{PK}$. Pick a random hash key τ for $\mathcal{H}$. The public-key is (τ, pk), the secret-key is (τ, sk).

$\mathsf{Enc}(pk, m)$**.** Pick $C \leftarrow_R \mathcal{V}$ together with its witness r that $C \in \mathcal{V}$. The session key $K = \mathcal{H}_\tau(\Lambda_{sk}(C)) \in \{0,1\}^l$ is computed as $K \leftarrow \mathcal{H}_\tau(\mathsf{Pub}(pk, C, r))$. The symmetric ciphertext is $\psi \leftarrow \mathsf{E}_K(m)$. Return the ciphertext (C, ψ).

$\mathsf{Dec}(sk, C)$**.** Reconstruct the key $K = \mathcal{H}_\tau(\Lambda_{sk}(C))$ as $K \leftarrow \mathcal{H}_\tau(\mathsf{Priv}(sk, C))$ and return $\{m, \perp\} \leftarrow \mathsf{D}_K(\psi)$.

Combining Theorems 4.1 and 4.3 gives us the following corollary.

Corollary 4.4 Assume HPS is ϵ_1-almost 1-universal with hard subset membership problem and with collision probability δ, that $\mathcal{HS}$ is a family of 4-wise independent hash functions with $\mathcal{H} : \mathcal{K} \to \{0,1\}^\ell$, and that SE is AE-OT secure. Then the encryption scheme above is secure in the sense of IND-CCA2. In particular,

$$
\mathsf{Adv}^{\mathsf{cca2}}_{\mathsf{PKE},t,Q}(k) \leq \mathsf{Adv}^{\mathsf{sm}}_{\mathsf{HPS},t}(k) + 2Q \cdot \mathsf{Adv}^{\mathsf{int\text{-}ot}}_{\mathsf{SE},t}(k) + \mathsf{Adv}^{\mathsf{ind\text{-}ot}}_{\mathsf{SE},t}(k) + Q \cdot (\frac{3}{2} \cdot \frac{2^\ell}{\sqrt{|\mathcal{K}|}} + 3\epsilon_1 + \delta) \; .
$$

5 Instantiations from the DDH Assumption

In this section we discuss two practical instantations of our randomness extraction framework whose security is based on the DDH assumption. A concrete instantiation from the QR assumption can be found in Appendix A.

5.1 The Decisional Diffie-Hellman (DDH) Assumption

A group scheme $\mathcal{GS}$ [5] specifies a sequence $(\mathcal{GR}_k)_{k\in\mathbb{N}}$ of group descriptions. For every value of a security parameter $k \in \mathbb{N}$, the pair $\mathcal{GR}_k = (\mathbb{G}_k, p_k)$ specifies a cyclic (multiplicative) group $\mathbb{G}_k$ of prime order p_k. Henceforth, for notational convenience, we tend to drop the index k. We assume the existence of an efficient sampling algorithm $x \leftarrow_R \mathbb{G}$ and an efficient membership algorithm. We define the ddh-advantage of an adversary B as

$$\mathsf{Adv}^{\mathrm{ddh}}_{\mathcal{GS},\mathsf{B}}(k) \stackrel{\mathrm{def}}{=} \left|\Pr[\mathsf{B}(g_1, g_2, g_1^r, g_2^r) = 1] - \Pr[\mathsf{B}(g_1, g_2, g_1^r, g_2^{\tilde{r}}) = 1]\right| ,$$

where $g_1,\ g_2 \leftarrow_R \mathbb{G}$, $r \leftarrow_R \mathbb{Z}_p$, $\tilde{r} \leftarrow_R \mathbb{Z}_p \setminus \{r\}$. We say that the DDH problem is hard in $\mathcal{GS}$ if the advantage function $\mathsf{Adv}^{\mathrm{ddh}}_{\mathcal{GS},\mathsf{B}}(k)$ is a negligible function in k for all probabilistic PTA B.

5.2 Variant 1: the Scheme HE_1

THE 1-UNIVERSAL HASH PROOF SYSTEM. We recall a 1-universal HPS by Cramer and Shoup [4], whose hard subset membership problem is based on the DDH assumption. Let $\mathcal{GS}$ be a group scheme where $\mathcal{GR}_k$ specifies $(\mathbb{G}, p)$ and let g_1, g_2 be two independent generators of $\mathbb{G}$. Define $\mathcal{C} = \mathbb{G}^2$ and $\mathcal{V} = \{(g_1^r, g_2^r) \subset \mathbb{G}^2 \ :\ r \in \mathbb{Z}_p\}$. The value $r \in \mathbb{Z}_p$ is a witness of $C \in \mathcal{V}$. The trapdoor generator Param picks a uniform trapdoor $\omega \in \mathbb{Z}_p$ and computes $g_2 = g_1^\omega$. Note that using trapdoor ω, algorithm Decide can efficienctly perform subset membership tests for $C = (c_1, c_2) \in \mathcal{C}$ by checking whether $c_1^\omega = c_2$.

Let $\mathcal{SK} = \mathbb{Z}_p^2$, $\mathcal{PK} = \mathbb{G}$, and $\mathcal{K} = \mathbb{G}$. For $sk = (x_1, x_2) \in \mathbb{Z}_p^2$, define $\mu(sk) = X = g_1^{x_1} g_2^{x_2}$. This defines the output of $\mathsf{Param}(1^k)$. For $C = (c_1, c_2) \in \mathcal{C}$ define

$$\Lambda_{sk}(C) := c_1^{x_1} c_2^{x_2}\ . \tag{8}$$

This defines $\mathsf{Priv}(sk, C)$. Given $pk = \mu(sk)$, $C \in \mathcal{V}$ and a witness $r \in \mathbb{Z}_p$ such that $C = (c_1, c_2) = (g_1^r, g_2^r)$ public evaluation $\mathsf{Pub}(pk, C, r)$ computes $K = \Lambda_{sk}(C)$ as

$$K = X^r\ .$$

Correctness follows by (8) and the definition of μ. This completes the description of HPS. Clearly, under the DDH assumption, the subset membership problem is hard in HPS. Moreover, this HPS is known to be (perfect) 1-universal [4]:

Lemma 5.1 The above HPS is perfect 1-universal (so $\epsilon_1 = 0$) with collision probability $\delta = 1/p$.

Proof: For perfect 1-universality, it suffices to show that given the public key X and any pair $(C, K) \in (\mathcal{C} \setminus \mathcal{V}) \times \mathcal{K}$, there exists exactly one secret key sk such that $\mu(sk) = X$ and $\Lambda_{sk}(C) = K$. Let $\omega \in \mathbb{Z}_p^*$ be such that $g_2 = g_1^\omega$, write $C = (g_1^r, g_2^s)$ for $r \neq s$ and consider a possible secret key $sk = (x_1, x_2) \in \mathbb{Z}_p^2$. Then we simultaneously need that $\mu(sk) = g_1^{x_1+\omega x_2} = X = g^x$ (for some $x \in \mathbb{Z}_p$) and $\Lambda_{sk}(C) = g_1^{rx_1+s\omega x_2} = K = g_1^y$ (for some $y \in \mathbb{Z}_p$). Then, using linear algebra, x_1 and x_2 follow uniquely from r, s, x, y and ω provided that the relevant determinant $(s-r)\omega \neq 0$. This is guaranteed here since $r \neq s$ and $\omega \neq 0$.

To verify the bound on the collision probability δ it suffices —due to symmetry— to determine for any distinct pair $(C, C^*) \in (\mathcal{C} \setminus \mathcal{V})^2$ the probability $\Pr_{sk}[\Lambda_{sk}(C) = \Lambda_{sk}(C^*)]$. In other words, for $(r,s) \neq (r',s')$ (with $r \neq s$ and $r' \neq s'$, but that is irrelevant here) we have that

$$\begin{aligned}\delta &= \Pr_{x_1,x_2 \leftarrow_R \mathbb{Z}_p}[g_1^{rx_1+x_2\omega s} = g_1^{r'x_1+x_2\omega s'}] \\ &= \Pr_{x_1,x_2 \leftarrow_R \mathbb{Z}_p}[rx_1 + x_2\omega s = r'x_1 + x_2\omega s'] \\ &= 1/p \ .\end{aligned}$$

(For the last step, use that if $r \neq r'$ for any x_2 only one x_1 will "work"; if $r = r'$ then necessarily $s \neq s'$ and for any x_1 there is a unique x_2 to satisfy the equation). ▮

THE HYBRID ENCRYPTION SCHEME HE_1. For our hybrid encryption scheme we make the following assumptions.

- Let $\mathcal{GS}$ be a group scheme where $\mathcal{GR}_k$ specifies $(\mathbb{G}, p)$ and the DDH assumption holds;
- Let $\mathcal{HS}$ be a family $\mathcal{H}_k : \mathbb{G} \to \{0,1\}^{\ell(k)}$ of 4-wise independent hash functions with $\lg p \geq 4\ell(k)$;
- Let $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ be a $\mathsf{AE\text{-}OT}$ secure symmetric scheme with key-space $\{0,1\}^{\ell(k)}$.

Applying the transformation from Theorem 4.3 one obtains an ϵ-almost 2-universal hash proof system with $\epsilon \leq 2 \cdot 2^{-\ell(k)}$ (using Lemma 5.1 and $|\mathbb{G}| = p \geq 2^{4\ell(k)}$). The resulting hybrid encryption scheme is depicted in Figure 1. Corollary 4.4 (in conjuction with Lemma 5.1) can be used to bound an adversary's $\mathsf{IND\text{-}CCA2}$ advantage.

Theorem 5.2 Let $\mathcal{GS} = (\mathbb{G}, p)$ be a group scheme where the DDH problem is hard, let $\mathcal{H}$ be a family of 4-wise independent hash functions from $\mathbb{G}$ to $\{0,1\}^{\ell(k)}$ with $\lg p \geq 4\ell(k)$, and let SE be a symmetric encryption that is secure in the sense of $\mathsf{AE\text{-}OT}$. Then HE_1 is secure in the sense of $\mathsf{IND\text{-}CCA2}$. In particular,

$$\mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{HE}_1,t,Q}(k) \leq \mathsf{Adv}^{\mathrm{ddh}}_{\mathcal{GS},t}(k) + 2Q \cdot \mathsf{Adv}^{\mathrm{int\text{-}ot}}_{\mathsf{SE},t}(k) + \mathsf{Adv}^{\mathrm{ind\text{-}ot}}_{\mathsf{SE},t}(k) + \frac{2Q}{2^{\ell(k)}} \ .$$

Proof: The only difference between the statement above and a direct application of Corollary 4.4 is the way we bound the loss due to the 1-HPS to 2-HPS transformation:

$$\frac{3}{2} \cdot \frac{2^{\ell}}{\sqrt{|\mathcal{K}|}} + 3\epsilon_1 + \delta = \frac{3}{2} \cdot \frac{2^{\ell}}{2^{2\ell}} + \frac{1}{2^{4\ell}} \leq 2^{-\ell+1} \ ,$$

where we used that $|\mathcal{K}| = |\mathbb{G}| = p \geq 2^{4\ell}$ and (by Lemma 5.1) $\epsilon_1 = 0$ and $\delta = 1/p$. ▮

In terms of concrete security, Theorem 5.2 requires the image $\{0,1\}^{\ell(k)}$ of $\mathcal{H}$ to be sufficiently small, i.e., $\ell(k) \leq \frac{1}{4}\lg p$. For a symmetric cipher with $\ell(k) = k = 80$ bits keys we are forced to use groups of order $\lg p = 4k = 320$ bits. For some specific groups such as elliptic curves this can be a drawback since there one typically works with groups of order $\lg p = 2k = 160$ bits.

RELATION TO DAMGÅRD'S ELGAMAL. In HE_1, invalid ciphertexts of the form $c_1^{\omega} \neq c_2$ are reject implicitly by authenticity properties of the symmetric cipher. Similar to [5], a variant of this scheme, $\mathsf{HE}_1^{\mathsf{er}} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$, in which such invalid ciphertexts get explicitly rejected is given in Figure 2. The scheme is slightly simplified compared to a direct explicit version that adds the trapdoor to the secret key; the simplification can be justified using the techniques of Lemma 5.1.

We remark that, interestingly, Damgård's encryption scheme [6] (also known as Damgård's ElGamal) is a special case of $\mathsf{HE}_1^{\mathsf{er}}$ where the hash function $\mathcal{H}$ is the identity function (or an

$\mathsf{Kg}(1^k)$	$\mathsf{Enc}(pk, m)$	$\mathsf{Dec}(sk, C)$
$x_1, x_2 \leftarrow_R \mathbb{Z}_p$; $X \leftarrow g_1^{x_1} g_2^{x_2}$	$r \leftarrow_R \mathbb{Z}_p^*$; $c_1 \leftarrow g_1^r$; $c_2 \leftarrow g_2^r$	Parse C as (c_1, c_2, ψ)
Pick random key τ for $\mathcal{H}$	$K \leftarrow \mathcal{H}_\tau(X^r) \in \{0,1\}^\ell$	$K \leftarrow \mathcal{H}_\tau(c_1^{x_1} c_2^{x_2})$
$pk \leftarrow (X, \tau)$; $sk \leftarrow (x_1, x_2)$	$\psi \leftarrow \mathsf{E}_K(m)$	Return $\{m, \perp\} \leftarrow \mathsf{D}_K(\psi)$
Return (sk, pk)	Return $C = (c_1, c_2, \psi)$	

Figure 1: Hybrid encryption scheme $\mathsf{HE}_1 = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$.

$\mathsf{Kg}(1^k)$	$\mathsf{Enc}(pk, m)$	$\mathsf{Dec}(sk, C)$
$\omega, x \leftarrow_R \mathbb{Z}_p$; $g_2 \leftarrow g_1^\omega$; $X \leftarrow g_1^x$	$r \leftarrow_R \mathbb{Z}_p^*$; $c_1 \leftarrow g_1^r$; $c_2 \leftarrow g_2^r$	Parse C as (c_1, c_2, ψ)
Pick random key τ for $\mathcal{H}$	$K \leftarrow \mathcal{H}_\tau(X^r) \in \{0,1\}^\ell$	if $c_1^\omega \neq c_2$ return $\perp$
$pk \leftarrow (g_2, X, \tau)$; $sk \leftarrow (x, \omega)$	$\psi \leftarrow \mathsf{E}_K(m)$	$K \leftarrow \mathcal{H}_\tau(c_1^x)$
Return (sk, pk)	Return $C = (c_1, c_2, \psi)$	Return $\{m, \perp\} \leftarrow \mathsf{D}_K(\psi)$

Figure 2: Hybrid encryption scheme $\mathsf{HE}_1^{\mathsf{er}} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ with explicit rejection.

easy-to-invert, canonical embedding of the group into, say, the set of bitstrings) and SE is "any easy to invert group operation" [6], for example the one-time pad with $\mathsf{E}_K(m) = K \oplus m$. In his paper, Damgård proved $\mathsf{IND\text{-}CCA1}$ security of his scheme under the DDH assumption and the *knowledge of exponent* assumption in $\mathcal{GS}$.[4] Our schemes $\mathsf{HE}_1^{\mathsf{er}}$ and HE_1 can therefore be viewed as hybrid versions of Damgård's ElGamal scheme, that can be proved secure under the DDH assumption.

5.3 Variant 2: the Scheme HE_2

The 1-universal hash proof system. We now give an alternative (and new) 1-universal hash proof system from the DDH assumption. Keep $\mathcal{C}$ and $\mathcal{V}$ as before. Define $\mathcal{SK} = \mathbb{Z}_p^4$, $\mathcal{PK} = \mathbb{G}^2$, and $\mathcal{K} = \mathbb{G}^2$. For $sk = (x_1, x_2, \hat{x}_1, \hat{x}_2) \in \mathbb{Z}^4$, define $\mu(sk) = (X, \hat{X}) = (g_1^{x_1} g_2^{x_2}, g_1^{\hat{x}_1} g_2^{\hat{x}_2})$. For $C = (c_1, c_2) \in \mathcal{C}$ define

$$\Lambda_{sk}(C) := (c_1^{x_1} c_2^{x_2}, c_1^{\hat{x}_1} c_2^{\hat{x}_2}) .$$

This also defines $\mathsf{Priv}(sk, C)$. Given $pk = \mu(sk)$, $C \in \mathcal{V}$ and a witness $r \in \mathbb{Z}_p$ such that $C = (c_1, c_2) = (g_1^r, g_2^r)$, public evaluation $\mathsf{Pub}(pk, C, r)$ computes $K = \Lambda_{sk}(C)$ as

$$K = (X^r, \hat{X}^r) .$$

Similar to Lemma 5.1 we can prove the following.

Lemma 5.3 The above HPS is perfect 1-universal ($\epsilon_1 = 0$) with collision probability $\delta = 1/p^2$.

The scheme HE_2. For our second hybrid encryption scheme HE_2 we make the same assumption as for HE_1, with the difference that $\mathcal{HS}$ is now a family $\mathcal{H}_k : \mathbb{G}^2 \to \{0,1\}^{\ell(k)}$ of 4-wise independent hash functions with $\lg p \geq 2\ell(k)$. Applying the transformation from Theorem 4.3 one obtains an ϵ-almost 2-universal hash proof system with $\epsilon \leq 2 \cdot 2^{-\ell(k)}$ (using Lemma 5.1 and $\lg |\mathcal{K}| =$

[4] To be more precise, Damgård only formally proved one-way ($\mathsf{OW\text{-}CCA1}$) security of his scheme, provided that the original ElGamal scheme is $\mathsf{OW\text{-}CPA}$ secure. But he also remarks that his proof can be reformulated to prove $\mathsf{IND\text{-}CCA1}$ security, provided that ElGamal itself is $\mathsf{IND\text{-}CPA}$ secure. $\mathsf{IND\text{-}CPA}$ security of ElGamal under the DDH assumption was only formally proved later [27].

$\mathsf{Kg}(1^k)$	$\mathsf{Enc}(pk, m)$	$\mathsf{Dec}(sk, C)$
$x_1, x_2, \hat{x}_1, \hat{x}_2 \leftarrow_R \mathbb{Z}_p$	$r \leftarrow_R \mathbb{Z}_p^*$; $c_1 \leftarrow g_1^r$; $c_2 \leftarrow g_2^r$	Parse C as (c_1, c_2, ψ)
$X \leftarrow g_1^{x_1} g_2^{x_2}$; $\hat{X} \leftarrow g_1^{\hat{x}_1} g_2^{\hat{x}_2}$	$K \leftarrow \mathcal{H}_\tau(X^r, \hat{X}^r) \in \{0,1\}^\ell$	$K \leftarrow \mathcal{H}_\tau(c_1^{x_1} c_2^{x_2}, c_1^{\hat{x}_1} c_2^{\hat{x}_2})$
Pick random key τ for $\mathcal{H}$	$\psi \leftarrow \mathsf{E}_K(m)$	Return $\{m, \perp\} \leftarrow \mathsf{D}_K(\psi)$
$pk \leftarrow (X, \hat{X}, \tau)$	Return $C = (c_1, c_2, \psi)$	
$sk \leftarrow (x_1, x_2, \hat{x}_1, \hat{x}_2)$		
Return (sk, pk)		

Figure 3: Hybrid encryption scheme $\mathsf{HE}_2 = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$.

$\lg|\mathbb{G}^2| = 2\lg p \geq 4\ell(k))$. The resulting hybrid encryption scheme is depicted in Figure 3. This time Corollary 4.4 (in conjuction with Lemma 5.3) leads to the following.

Theorem 5.4 Let $\mathcal{GS} = (\mathbb{G}, p)$ be a group scheme where the DDH problem is hard, let $\mathcal{H}$ be a family of 4-wise independent hash functions from $\mathbb{G}^2$ to $\{0,1\}^{\ell(k)}$ with $\lg p \geq 2\ell(k)$, and let SE be a symmetric encryption that is secure in the sense of AE-OT. Then HE_2 is secure in the sense of IND-CCA2. In particular,

$$\mathsf{Adv}^{\mathrm{cca2}}_{\mathsf{HE}_2,t,Q}(k) \leq \mathsf{Adv}^{\mathrm{ddh}}_{\mathcal{GS},t}(k) + 2Q \cdot \mathsf{Adv}^{\mathrm{int\text{-}ot}}_{\mathsf{SE},t}(k) + \mathsf{Adv}^{\mathrm{ind\text{-}ot}}_{\mathsf{SE},t}(k) + \frac{2Q}{2^{\ell(k)}} .$$

Note that HE_2 now only has the restriction $\lg p \geq 2\ell(k)$ which nicely fits with the typical choice of $\ell(k) = k$ and $\lg p = 2k$. So one is free to use any cryptographic group, in particular also elliptic curve groups.

Similar to $\mathsf{HE}_1^{\mathsf{er}}$, the variant $\mathsf{HE}_2^{\mathsf{er}}$ with explicit rejection can again be proven equivalent. In the explicit rejection variant, $\mathsf{HE}_2^{\mathsf{er}}$, the public-key contains the group elements $g_2 = g_1^\omega$, $X = g_1^x$, and $\hat{X} = g_1^{\hat{x}}$), and decryption first checks if $c_1^\omega = c_2$ and then computes $K = \mathcal{H}_\tau(c_1^x, c_1^{\hat{x}})$.

Relation to a scheme by Kurosawa and Desmedt. We remark that, interestingly, the scheme HE_2 is quite similar to the one by Kurosawa and Desmedt [19]. The only difference is that encryption in the latter defines the key as $K = X^{rt} \cdot \hat{X}^r \in \mathbb{G}$, where $t = \mathsf{T}_\tau(c_1, c_2)$ is the output of a target collision-resistant hash function $\mathsf{T}_\tau : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{Z}_p$.

6 Efficiency Considerations

In this section we compare the efficiency of HE_1/HE_2 and their explicit rejection variants $\mathsf{HE}_1^{\mathsf{er}}$/$\mathsf{HE}_2^{\mathsf{er}}$ with the reference scheme by Kurosawa and Desmedt [19] and its variants [13, 17].

The drawback of HE_1 is that, in terms of concrete security, Theorem 5.2 requires the image $\{0,1\}^\ell$ of $\mathcal{H}$ to be sufficiently small, i.e., $\ell \leq \frac{1}{4}\lg p$. Consequently, for a symmetric cipher with $\ell = k = 80$ bits keys we are forced to use groups of order $\lg p \geq 4k = 320$ bits. For some specific groups such as elliptic curves this can be a drawback since there one typically works with groups of order $\lg = 2k = 160$ bits. However, for other more traditional groups such as prime subgroups of $\mathbb{Z}_q^*$ one sometimes takes a subgroup of order already satisfying the requirement $\lg p \geq 4k$. The scheme HE_2 overcomes this restriction at the cost of an additional exponentiation in the encryption algorithm.

Table 1 summarizes the efficiency of the schemes KD [19], $\mathsf{HE}_1^{\mathsf{er}}$, and $\mathsf{HE}_2^{\mathsf{er}}$. (A comparison of the explicit rejection variants seems more meaningful.) It is clear that when groups of similar size are used, our new scheme $\mathsf{HE}_1^{\mathsf{er}}$ will be the most efficient. But, as detailed above, typically $\mathsf{HE}_1^{\mathsf{er}}$

will have to work in a larger (sub)group. Even when underlying operations such as multiplication and squaring remain the same, the increased exponent length will make this scheme noticeably slower than the other two options.

For any given group, it is hard to tell in advance whether $\mathsf{HE}_2^{\mathsf{er}}$ or KD will be fastest. For encryption, they differ only in one point: the key derivation. In $\mathsf{HE}_2^{\mathsf{er}}$ the symmetric key is computed as $K = \mathcal{H}_\tau(X^r, \hat{X}^r) \in \{0,1\}^k$ where $\mathcal{H}_\tau$ is a 4-wise independent hash function, in KD it is computed as $K = X^{rt} \cdot \hat{X}^r \in \mathbb{G}$, where $t = \mathsf{T}_\tau(c_1, c_2)$, for a target collision resistant hash function T (plus possibly the application of a key-derivation function KDF to represent the group element K as a bit string suitable for symmetric key, cf. Appendix B). If one would ignore the hashing, we see that we need to compute two single exponentiations (distinct bases, same exponent) in $\mathsf{HE}_2^{\mathsf{er}}$ versus a double exponentiation in KD. It is well known that in most scenarios a double exponentiation costs significantly less than two separate single exponentiations. In practice this is mainly due to the possibility to combine the squarings from both components of the double exponentiation, thus saving $\lg p$ squarings (when compared to two single exponentiations) and, to a lesser degree, to the ability to encode the two exponents simultaneously in such a way that the weight is less than twice that of a single encoded exponent, thus saving on multiplications. This all benefits KD. However, it should be noted that in certain scenarios the advantage is less pronounced, e.g., when squaring is for free or when precomputation on the public key (including group elements X and $\hat{X}$) should be taken into account. We remark that the decryption algorithms of KD and $\mathsf{HE}_2^{\mathsf{er}}$ have roughly the same efficiency: KD uses two exponentiations (to compute c_1^ω and c_1^x), and $\mathsf{HE}_2^{\mathsf{er}}$ three (to compute c_1^ω, c_1^x, and $c_1^{\hat{x}}$)). It is well known that exponentiations with respect to the same basis can be computed quite efficiently in one go. The additional cost of having a third exponent in our case does therefore not incur too much of a performance penalty.

Indeed, the main computational advantage of our scheme lies in the much simpler hash that is required to attain provable security. A 4-wise independent hash function is a combinatorial object that can be implemented with three multiplications (typically in a field of size $\approx p$, not in $\mathbb{G}$ itself). On the other hand, a target collision resistant hash function is a computational object. Bootstrapping such a hash function from the presumed hardness of the DDH problem will not be cheap. It will most likely cost at least the equivalent of one exponentiation.[5] In that case $\mathsf{HE}_2^{\mathsf{er}}$ will be faster than KD, both for encryption and decryption. Another important advantage of $\mathsf{HE}_2^{\mathsf{er}}$ is that in encryption the computation of c_1, c_2, and the key $(X^r, \hat{X}^r)$ can be done in parallel, whereas in KD the computation of the key $X^{r \cdot \mathsf{T}(c_1,c_2)}\hat{X}^r$ can only be done *after* the values c_1 and c_2 are available.

Acknowledgements

We thank Ronald Cramer for interesting discussions. We are furthermore grateful to Victor Shoup for pointing out the scheme from Section 5.3. We thank Kenny Paterson, Steven Galbraith and James Birkett for useful feedback, prompting the comparison in Section 6.

[5] To the best of our knowledge, the most efficient construction of a (target) collision resistant hash function which is provably secure in groups with hard DDH problem is the function $\mathcal{H}_\tau(x_1, x_2) := A_1^{x_1} A_2^{x_2} \in \mathbb{G}$, where $\tau = (A_1, A_2) \in \mathbb{G}^2$).

Scheme	Assumption	Encryption #[multi/sequential,single]-exp	Decryption	Ciphertext Size	Key-size Public	Secret	Restriction on $p = ord(\mathbb{G})$
KD	DDH & TCR	$[1,2]$+tcr	$[1,0]$+tcr	$2\lvert\mathbb{G}\rvert+\lvert\psi\rvert$	$4\lvert\mathbb{G}\rvert+\lvert\tau_{\mathrm{tcr}}\rvert$	$4\lvert\mathbb{Z}_p\rvert$	$\lg p \geq 2\ell(k)$
$\mathsf{HE}_1^{\mathsf{er}}$	DDH	$[0,3]$+4wh	$[1,0]$+4wh	$2\lvert\mathbb{G}\rvert+\lvert\psi\rvert$	$3\lvert\mathbb{G}\rvert+\lvert\tau_{4wh}\rvert$	$2\lvert\mathbb{Z}_p\rvert$	$\lg p \geq 4\ell(k)$
$\mathsf{HE}_2^{\mathsf{er}}$	DDH	$[0,4]$+4wh	$[1,0]$+4wh	$2\lvert\mathbb{G}\rvert+\lvert\psi\rvert$	$4\lvert\mathbb{G}\rvert+\lvert\tau_{4wh}\rvert$	$4\lvert\mathbb{Z}_p\rvert$	$\lg p \geq 2\ell(k)$

Table 1: Efficiency comparison for known CCA2-secure encryption schemes from the DDH assumption. All "symmetric" operations concerning the authenticated encryption scheme are ignored. The symbols "tcr" and "4wh" denote one application of a target collision-resistant hash function and 4-wise independent hash function, respectively.

References

[1] Mihir Bellare and Chanathip Namprempre. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In Tatsuaki Okamoto, editor, *ASIACRYPT 2000*, volume 1976 of *LNCS*, pages 531–545. Springer-Verlag, Berlin, Germany, December 2000. (Cited on page 5, 19, 20.)

[2] Olivier Chevassut, Pierre-Alain Fouque, Pierrick Gaudry, and David Pointcheval. Key derivation and randomness extraction. Cryptology ePrint Archive, Report 2005/061, 2005. `http://eprint.iacr.org/`. (Cited on page 3.)

[3] Ronald Cramer and Victor Shoup. A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In Hugo Krawczyk, editor, *CRYPTO'98*, volume 1462 of *LNCS*, pages 13–25. Springer-Verlag, Berlin, Germany, August 1998. (Cited on page 1, 2.)

[4] Ronald Cramer and Victor Shoup. Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In Lars R. Knudsen, editor, *EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 45–64. Springer-Verlag, Berlin, Germany, April / May 2002. (Cited on page 1, 2, 3, 4, 7, 10.)

[5] Ronald Cramer and Victor Shoup. Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal on Computing*, 33(1):167–226, 2003. (Cited on page 1, 2, 3, 5, 10, 11, 19, 20.)

[6] Ivan Damgård. Towards practical public key systems secure against chosen ciphertext attacks. In Joan Feigenbaum, editor, *CRYPTO'91*, volume 576 of *LNCS*, pages 445–456. Springer-Verlag, Berlin, Germany, August 1992. (Cited on page 2, 11, 12.)

[7] Yvo Desmedt and Duong Hieu Phan. A CCA secure hybrid Damgårds ElGamal encryption. In *ProvSec 2008*, volume 5324, pages ???–??? LNCS, 2008. (Cited on page 2.)

[8] Yevgeniy Dodis, Rosario Gennaro, Johan Håstad, Hugo Krawczyk, and Tal Rabin. Randomness extraction and key derivation using the CBC, cascade and HMAC modes. In Matthew Franklin, editor, *CRYPTO 2004*, volume 3152 of *LNCS*, pages 494–510. Springer-Verlag, Berlin, Germany, August 2004. (Cited on page 3.)

[9] Danny Dolev, Cynthia Dwork, and Moni Naor. Non-malleable cryptography. In *23rd ACM STOC*, pages 542–552. ACM Press, May 1991. (Cited on page 1.)

[10] Danny Dolev, Cynthia Dwork, and Moni Naor. Nonmalleable cryptography. *SIAM Journal on Computing*, 30(2):391–437, 2000. (Cited on page 3.)

[11] Taher ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In G. R. Blakley and David Chaum, editors, *CRYPTO'84*, volume 196 of *LNCS*, pages 10–18. Springer-Verlag, Berlin, Germany, August 1985. (Cited on page 2.)

[12] Rosario Gennaro, Hugo Krawczyk, and Tal Rabin. Secure Hashed Diffie-Hellman over non-DDH groups. In Christian Cachin and Jan Camenisch, editors, *EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 361–381. Springer-Verlag, Berlin, Germany, May 2004. (Cited on page 3.)

[13] Rosario Gennaro and Victor Shoup. A note on an encryption scheme of Kurosawa and Desmedt. Cryptology ePrint Archive, Report 2004/194, 2004. `http://eprint.iacr.org/`. (Cited on page 7, 13.)

[14] Kristian Gjøsteen. A new security proof for Damgård's ElGamal. In David Pointcheval, editor, *CT-RSA 2006*, volume 3860 of *LNCS*, pages 150–158. Springer-Verlag, Berlin, Germany, February 2006. (Cited on page 2.)

[15] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. (Cited on page 2, 5.)

[16] D. Hofheinz, J. Herranz, and E. Kiltz. The Kurosawa-Desmedt key encapsulation is not chosen-ciphertext secure. Cryptology ePrint Archive, Report 2006/207, 2006. `http://eprint.iacr.org/`. (Cited on page 7.)

[17] Dennis Hofheinz and Eike Kiltz. Secure hybrid encryption from weakened key encapsulation. In Alfred Menezes, editor, *CRYPTO 2007*, volume 4622 of *LNCS*, pages 553–571. Springer-Verlag, Berlin, Germany, August 2007. (Cited on page 4, 7, 8, 13.)

[18] Jonathan Katz and Moti Yung. Unforgeable encryption and chosen ciphertext secure modes of operation. In Bruce Schneier, editor, *FSE 2000*, volume 1978 of *LNCS*, pages 284–299. Springer-Verlag, Berlin, Germany, April 2000. (Cited on page 20.)

[19] Kaoru Kurosawa and Yvo Desmedt. A new paradigm of hybrid encryption scheme. In Matthew Franklin, editor, *CRYPTO 2004*, volume 3152 of *LNCS*, pages 426–442. Springer-Verlag, Berlin, Germany, August 2004. (Cited on page 1, 2, 4, 7, 13.)

[20] Helger Lipmaa. On CCA1-Security of Elgamal and Damgård cryptosystems. Cryptology ePrint Archive, Report 2008/234, 2008. `http://eprint.iacr.org/`. (Cited on page 2.)

[21] Moni Naor. On cryptographic assumptions and challenges (invited talk). In Dan Boneh, editor, *CRYPTO 2003*, volume 2729 of *LNCS*, pages 96–109. Springer-Verlag, Berlin, Germany, August 2003. (Cited on page 2.)

[22] Moni Naor and Moti Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *22nd ACM STOC*. ACM Press, May 1990. (Cited on page 1.)

[23] Duong Hieu Phan and David Pointcheval. About the security of ciphers (semantic security and pseudo-random permutations). In Helena Handschuh and Anwar Hasan, editors, *SAC 2004*, volume 3357 of *LNCS*, pages 182–197. Springer-Verlag, Berlin, Germany, August 2004. (Cited on page 20.)

[24] Charles Rackoff and Daniel R. Simon. Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. In Joan Feigenbaum, editor, *CRYPTO'91*, volume 576 of *LNCS*, pages 433–444. Springer-Verlag, Berlin, Germany, August 1992. (Cited on page 1, 3.)

[25] Hovav Shacham. A cramer-shoup encryption scheme from the linear assumption and from progressively weaker linear variants. Cryptology ePrint Archive, Report 2007/074, 2007. `http://eprint.iacr.org/`. (Cited on page 7.)

[26] Victor Shoup. Using hash functions as a hedge against chosen ciphertext attack. In Bart Preneel, editor, *EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 275–288. Springer-Verlag, Berlin, Germany, May 2000. (Cited on page 1, 2.)

[27] Yiannis Tsiounis and Moti Yung. On the security of ElGamal based encryption. In Hideki Imai and Yuliang Zheng, editors, *PKC'98*, volume 1431 of *LNCS*, pages 117–134. Springer-Verlag, Berlin, Germany, February 1998. (Cited on page 12.)

[28] Mark N. Wegman and Larry Carter. New hash functions and their use in authentication and set equality. *Journal of Computer and System Sciences*, 22:265–279, 1981. (Cited on page 19.)

[29] J. Wu and D.R. Stinson. On the security of the ElGamal encryption scheme and Damgard's variant. Cryptology ePrint Archive, Report 2008/200, 2008. `http://eprint.iacr.org/`. (Cited on page 2.)

A A universal HPS from the QR assumption

Let $N = pq$ be an RSA modulus, where $p = 2P+1$ and $q = 2Q+1$, for two primes P, Q. Let L_N denote the subgroup of elements in $\mathbb{Z}_N^*$ with Jacobi symbol 1, and let $\mathbb{QR}_N$ denote the unique (cyclic) subgroup of $\mathbb{Z}_N^*$ of order PQ (so in particular $\mathbb{QR}_N \subset L_N$). Let g be a generator of $\mathbb{QR}_N$. We assume the existence of an RSA instance generator RSAgen that generates the above elements. The quadratic residue (QR) assumption states that distinguishing a random element from $\mathbb{QR}_N$ from a random element from L_N is computationally infeasible.

Define $\mathcal{C} = \mathbb{Z}_N^*$ and $\mathcal{V} = \mathbb{QR}_N = \{g^r \ : \ r \in \mathbb{Z}_{PQ}\}$. The value $r \in \mathbb{Z}$ is a witness of $C \in \mathcal{V}$. (Note that it is possible to sample an almost uniform element from $\mathcal{V}$ together with a witness by first picking $r \in \mathbb{Z}_{\lfloor N/4 \rfloor}$ and defining $C = g^r$.) Define $\mathcal{SK} = \mathbb{Z}_{2PQ}^\ell$, $\mathcal{PK} = \mathbb{QR}_N^\ell$, and $\mathcal{K} = \{0,1\}^k$. For $sk = (x_1, \ldots, x_\ell) \in \mathbb{Z}_{2PQ}^\ell$, define $\mu(sk) = (X_1, \ldots, X_\ell) = (g^{x_1}, \ldots, g^{x_\ell})$. (Note that X_i does not reveal whether $0 \le x_i < PQ$ or $PQ \le x_i < 2PQ$.)

We assume a family of hash functions $\mathcal{HS}$ with $\mathcal{H} : (L_N)^\ell \to \{0,1\}^k$. (In practice one can use $\mathcal{H} : \mathbb{Z}_n^\ell \to \{0,1\}^k$, bearing in mind that perfect 4-wise independence of the latter only gives rise to almost 4-wise independence of the former.) For $C \in \mathcal{C}$ define

$$\Lambda_{sk}(C) := \mathcal{H}_\tau(C^{x_1}, \ldots, C^{x_\ell}) \ .$$

This defines $\mathsf{Priv}(sk, C)$. Given $pk = \mu(sk)$, $C \in \mathcal{V}$ and a witness $r \in \mathbb{Z}_{PQ}$ such that $C = g^r$, public evaluation $\mathsf{Pub}(pk, C, r)$ computes $K = \Lambda_{sk}(C)$ as

$$K = \mathcal{H}_\tau(X_1^r, \ldots, X_\ell^r) \ .$$

This completes the description of HPS. Under the QR assumption, the subset membership problem is hard in HPS. For $C \in \mathcal{C} \setminus \mathcal{V}$, given $pk = \mu(sk)$, each of the C^{x_i} contains exactly one

$\mathsf{Kg}(1^k)$	$\mathsf{Enc}(pk, m)$	$\mathsf{Dec}(sk, C)$
$(N, P, Q, g) \leftarrow_R \mathsf{RSAgen}(1^k)$	$r \leftarrow_R \mathbb{Z}_{\lfloor N/4 \rfloor}$	Parse C as (c, ψ)
For $i = 1$ to $4k$ do	$c \leftarrow g^r$	$K \leftarrow \mathcal{H}_\tau(c^{x_1}, \ldots, c^{x_{4k}})$
$x_i \leftarrow_R \mathbb{Z}_{2PQ}$; $X_i \leftarrow g^x$	$K \leftarrow \mathcal{H}_\tau(X_1^r, \ldots, X_{4k}^r)$	Return $\{m, \perp\} \leftarrow \mathsf{D}_K(\psi)$
Pick random key τ for $\mathcal{H}$	$\psi \leftarrow \mathsf{E}_K(m)$	
$pk \leftarrow (N, g, (X_i), \tau)$; $sk \leftarrow ((x_i))$	Return $C = (c, \psi)$	
Return (sk, pk)		

Figure 4: The hybrid encryption scheme from the QR assumption.

bit of min entropy such that $H_\infty((C^{x_1}, \ldots, C^{x_\ell}) \mid (pk, C)) = \ell$. Therefore, if $\mathcal{HS}$ is a family of 2-wise independent hash functions and $\ell \geq 2k$, then HPS is 1-universal. An application of Theorem 4.3 immediately yields a 2-universal HPS. However, since the above HPS already contains a family of universal hash functions, we may as well obtain a direct construction of a 2-universal HPS. Concretely, we can prove the following:

Lemma A.1 Assume the QR assumption holds, $\mathcal{HS}$ is a 4-wise independent hash function and $\ell \geq 4k$. Then HPS is a 2-universal HPS.

The resulting encryption scheme (which is depicted in Figure 4) has very compact ciphertexts but encryption/decryption are quite expensive since they require $\ell = 4k$ exponentiations in $\mathbb{Z}_N^*$. (Note that decryption can be sped up considerably compared to encryption by using CRT and multi-exponentiation techniques.)

B Authenticated symmetric encryption schemes

B.1 Security notions

Ciphertext Indistinguishability. Let $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ be a symmetric encryption scheme, and let $\mathsf{A} = (\mathsf{A}_1, \mathsf{A}_2)$ be an adversary. The advantage of A in breaking the ciphertext indistinguishability security of SE is:

$$\mathsf{Adv}_{\mathsf{SE},\mathsf{A}}^{\text{ind-ot}}(k) \stackrel{\text{def}}{=} \left| \Pr\left[b = b' : \begin{array}{l} K^* \leftarrow_R \mathcal{K}_{\mathsf{SE}}(k) \; ; \; (m_0, m_1, St) \leftarrow_R \mathsf{A}_1(1^k) \; ; \\ b \leftarrow_R \{0,1\} \; ; \; \psi^* \leftarrow_R \mathsf{E}_{K^*}(m_b) \; ; \; b' \leftarrow_R \mathsf{A}_2(1^k, St, \psi^*) \end{array} \right] - 1/2 \right|$$

The symmetric encryption scheme SE is one-time secure in the sense of *indistinguishability* ($\mathsf{IND\text{-}OT}$) if i for every adversary A with probabilistic PTA A_1 and A_2, the advantage $\mathsf{Adv}_{\mathsf{SE},A}^{\text{ind-ot}}(\cdot)$ is negligible.

Ciphertext Integrity. This captures the property that no efficient adversary can produce a new valid ciphertext after seeing the encryption of a single message. Let $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ be a symmetric encryption scheme, and let $\mathsf{A} = (\mathsf{A}_1, \mathsf{A}_2)$ be an algorithm.

$$\mathsf{Adv}_{\mathsf{SE},\mathsf{A}}^{\text{int-ot}}(k) \stackrel{\text{def}}{=} \Pr\left[\psi \neq \psi^* \wedge \mathsf{D}_{K^*}(\psi) \neq \perp : \begin{array}{l} K^* \leftarrow_R \mathcal{K}_{\mathsf{SE}}(k) \; ; \; (m, St) \leftarrow_R \mathsf{A}_1(1^k) \; ; \\ \psi^* \leftarrow \mathsf{E}_{K^*}(m) \; ; \; \psi \leftarrow_R \mathsf{A}_2(1^k, St, \psi^*) \end{array} \right]$$

The symmetric encryption scheme SE is one-time secure in the sense of *ciphertext integrity* ($\mathsf{INT\text{-}OT}$) if for every adversary A with probabilistic PTA A_1 and A_2, the advantage $\mathsf{Adv}_{\mathsf{SE},\mathsf{A}}^{\text{int-ot}}(\cdot)$ is negligible.

We also define weak ciphertext integrity ($\mathsf{WINT\text{-}OT}$) where in the above security experiment the adversary (in the second stage) never sees the ciphertext ψ^*. The corresponding advantage function is denoted as $\mathsf{Adv}^{\text{wint-ot}}_{\mathsf{SE},\mathsf{A}}$.

One-time Authenticated Encryption. A symmetric encryption scheme is secure in the sense of *one-time authenticated encryption* ($\mathsf{AE\text{-}OT}$) iff it is $\mathsf{IND\text{-}OT}$ and $\mathsf{INT\text{-}OT}$ secure. For the notion of *weak one-time authenticated encryption* ($\mathsf{WAE\text{-}OT}$) we only require it to be $\mathsf{IND\text{-}OT}$ and $\mathsf{WINT\text{-}OT}$ secure.

We now recall details of the encrypt-then-mac approach [1, 5] for constructing authenticated symmetric encryption.

B.2 Building blocks

Key Derivation Functions. A key-derivation function KDF is a family of functions $\mathsf{KDF}_k : \{0,1\}^\ell \to \{0,1\}^{2k}$. We assume its output on a random input is computationally indistinguishable from a random $2k$-bit string (pseudorandomness), captured by defining the pr-advantage of an adversary B_{kdf} as

$$\mathsf{Adv}^{\text{pr}}_{\mathsf{KDF},\mathsf{B}_{\text{kdf}}}(k) = |\Pr[\mathsf{B}_{\text{kdf}}(\mathsf{KDF}(K)) = 1] - \Pr[\mathsf{B}_{\text{kdf}}(X) = 1]|,$$

where $K \leftarrow_R \{0,1\}^\ell$ and $X \leftarrow_R \{0,1\}^{2k}$.

Message Authentication Codes. A message authentication code $\mathsf{MAC} = (\mathsf{Tag}, \mathsf{Vfy})$ with keys $mk \in \{0,1\}^k$ consists of a tag algorithm $\mathsf{Tag}_{mk}(m)$ and a verification algorithm $\mathsf{Vfy}_{mk}(\tau)$. For consistency we require that for all messages M, we have $\Pr[\mathsf{Vfy}_{mk}(M, \mathsf{Tag}_{mk}(M)) \neq \bot] = 1$, where the probability is taken over the choice of coins of all the algorithms in the expression above.

MAC needs to be *strongly unforgeable against one-time attacks* ($\mathsf{SUF\text{-}OT}$) captured by defining the suf-ot-advantage of an adversary B_{mac} as

$$\mathsf{Adv}^{\text{suf-ot}}_{\mathsf{MAC},\mathsf{B}_{mac}}(k) = \Pr[\mathsf{Vfy}_{mk}(m^*, \tau^*) \neq \bot \;:\; mk \leftarrow_R \{0,1\}^k \,;\, (M^*, \tau^*) \leftarrow_R \mathsf{B}_{mac}^{\mathsf{Tag}_{mk}(\cdot)}(1^k)] \,.$$

Above, oracle $\mathsf{Tag}_{mk}(\cdot)$ returns $\tau \leftarrow \mathsf{Tag}_{mk}(m)$ and A may only make one single query to oracle $\mathsf{Tag}_{mk}(\cdot)$. The target pair (m^*, τ^*) must be different from the pair (m, τ) obtained from $\mathsf{Tag}_{mk}(\cdot)$ (strong unforgeability).

We remark that efficient MACs satisfying the above definition can be constructed without any computational assumption (and secure against unbounded adversaries) using, e.g., almost strongly-universal hash families [28].

B.3 Construction of AE-OT and WAE-OT secure ciphers

Let $\mathsf{OTP} = (\tilde{\mathsf{E}}, \tilde{\mathsf{D}})$ be a symmetric encryption that inputs keys from $\{0,1\}^k$, let KDF a key-derivation function that outputs bitstrings of length $2k$, and let MAC be a MAC scheme with keys $mk \in \{0,1\}^k$. Using the "Encrypt-then-MAC" paradigm we can construct $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ that inputs keys $K \in \{0,1\}^\ell$ as follows.

$\mathsf{E}_K(m)$	$\mathsf{D}_K(\psi = (\psi', \tau))$
$(mk\|\|dk) \leftarrow \mathsf{KDF}(K)$, where $mk, dk \in \{0,1\}^k$	$(mk\|\|dk) \leftarrow \mathsf{KDF}(K)$
$\psi' \leftarrow \tilde{\mathsf{E}}_{dk}(m)$	If $\mathsf{Vfy}_{mk}(\psi', \tau) = \bot$ return $\bot$
$\tau \leftarrow \mathsf{Tag}_{mk}(\psi')$	$M \leftarrow \tilde{\mathsf{D}}_{dk}(\psi')$
Return $\psi = (\psi', \tau)$	Return M

Typically, a MAC tag (from a computationally secure MAC) has k bits, so the above construction generates ciphertexts of size $d(k) = |m| + k$. The following lemma [5, 18, 1] guarantees the AE scheme is one-time secure.

Lemma B.1 Assume OTP is $\mathsf{IND\text{-}OT}$, KDF is pseudorandom, and MAC is $\mathsf{SUF\text{-}OT}$. Then SE is $\mathsf{AE\text{-}OT}$. In particlar, we have

$$\mathsf{Adv}^{\text{ind-ot}}_{\mathsf{SE},t}(k) \leq \mathsf{Adv}^{\text{pr}}_{\mathsf{KDF},t}(k) + \mathsf{Adv}^{\text{ind-ot}}_{\mathsf{OTP},t}(k), \quad \mathsf{Adv}^{\text{int-ot}}_{\mathsf{SE},t}(k) \leq \mathsf{Adv}^{\text{pr}}_{\mathsf{KDF},t}(k) + \mathsf{Adv}^{\text{suf-ot}}_{\mathsf{MAC},t}(k) \ .$$

We remark that for authenticated encryption is a strictly stronger security notion than chosen-ciphertext security (using a separation example from [1]), whereas the latter is already sufficient for the KEM/DEM composition theorem [5] (i.e., a $\mathsf{IND\text{-}CCA2}$ secure KEM plus chosen-ciphertext secure symmetric encryption implies $\mathsf{IND\text{-}CCA2}$ secure PKE). On the other hand, there exists redundancy-free chosen-ciphertext secure symmetric encryption [23] (with $d(k) = |m|$) whereas redundancy-free authenticated encryption do not exist.

If we only require $\mathsf{WAE\text{-}OT}$ security, we can construct $\mathsf{SE} = (\mathsf{E}, \mathsf{D})$ without a MAC as follows.

$\mathsf{E}_K(m)$	$\mathsf{D}_K(\psi = (\psi', mk'))$
$(mk\|\|dk) \leftarrow \mathsf{KDF}(K)$, where $mk, dk \in \{0,1\}^k$	$(mk\|\|dk) \leftarrow \mathsf{KDF}(K)$
$\psi' \leftarrow \tilde{\mathsf{E}}_{dk}(m)$	If $mk \neq mk'$ return $\perp$
Return $\psi = (\psi', mk)$	Returm $m \leftarrow \tilde{\mathsf{D}}_{dk}(\psi')$

Lemma B.2 Assume OTP is $\mathsf{IND\text{-}OT}$ and KDF is pseudorandom. Then SE is $\mathsf{WAE\text{-}OT}$. In particlar, we have

$$\mathsf{Adv}^{\text{ind-ot}}_{\mathsf{SE},t}(k) \leq \mathsf{Adv}^{\text{pr}}_{\mathsf{KDF},t}(k) + \mathsf{Adv}^{\text{ind-ot}}_{\mathsf{OTP},t}(k), \quad \mathsf{Adv}^{\text{int-ot}}_{\mathsf{SE},t}(k) \leq \mathsf{Adv}^{\text{pr}}_{\mathsf{KDF},t}(k) \ .$$