Smooth Projective Hashing and Two-Message Oblivious Transfer

Shai Halevi
IBM Research

Yael Tauman Kalai*
Microsoft Research

March 30, 2007

Abstract

We present a general framework for constructing two-message oblivious transfer protocols using a modification of Cramer and Shoup's notion of smooth projective hashing (2002). This framework is an abstraction of the two-message oblivious transfer protocols of Naor and Pinkas (2001) and Aiello et al. (2001), whose security is based on the Decisional Diffie Hellman Assumption. In particular, we give two new oblivious transfer protocols. The security of one is based on the Quadratic Residuosity Assumption, and the security of the other is based on the N'th Residuosity Assumption. Our security guarantees are not simulation based, but are similar to the guarantees of the aforementioned two constructions. Compared to other applications of smooth projective hashing, in our context we must deal also with maliciously chosen parameters, which raises new technical difficulties.

We also improve on prior constructions of factoring-based smooth universal hashing, in that our constructions *do not require that the underlying RSA-composite is a product of safe primes.* In fact, we observe that the safe-prime requirement is unnecessary for many prior constructions. In particular, we observe that the factoring-based CCA secure encryption schemes due to Cramer-Shoup Gennaro-Lindell and Camenisch-Shoup remain secure even if the underlying RSA-composite is not a product of safe primes. (This holds for the schemes based on the Quadratic Residuosity Assumption as well as the ones based on the N'th Residuosity Assumption.)

1 Introduction

In [CS98], Cramer and Shoup introduced the first CCA secure encryption scheme, whose security is based on the Decisional Diffie Hellman (DDH) Assumption. They later presented an abstraction of this scheme based on a new notion that they called "smooth projective hashing" [CS02]. This abstraction yielded two new CCA secure encryption schemes; the security of one is based on the Quadratic Residuosity Assumption and the security of the other is based on the N'th Residuosity Assumption [Pa99].[1] This notion of smooth projective hashing was later used by Gennaro and Lindell [GL03] in the context of key generation from humanly memorizable passwords. That work abstracts and generalizes an earlier protocol for this problem [KOY01], whose security is based on the DDH Assumption.

*Much of this work was carried out while visiting IBM T.J. Watson Research, New York. Also supported in part by NSF CyberTrust grant CNS-0430450.

[1]The N'th Residuosity Assumption is also referred to in the literature as the Decisional Composite Residuosity Assumption and as Paillier's Assumption.

In this paper, we use smooth projective hashing to construct efficient two-message oblivious transfer protocols. Our work follows the same pattern, in that it abstracts and generalizes earlier protocols for this problem [NP01, AIR01] whose security is based on the DDH Assumption. Using smooth projective hashing in this context raises a new issue. Specifically, we must deal with the case that the hash family itself is chosen maliciously by the adversary. To this end, we add an extra requirement to the definition of smooth projective hashing. This issue did not arise in the previous two applications because these were either in the public key model or in the common reference string model.

We show that even with this additional requirement, we can still construct smooth projective hashing from any of the following assumptions: the DDH Assumption, the N'th Residuosity Assumption, and the Quadratic Residuosity Assumption. Moreover, for the last two constructions we can prove security even when the underlying RSA-composite is not a product of safe primes. We note that all previous factoring-based constructions of smooth projective hashing did rely on the assumption that the underlying RSA-composite is a product of safe primes.

1.1 Oblivious Transfer

Oblivious transfer is a protocol between a *sender*, holding two strings γ_0 and γ_1, and a *receiver* holding a choice bit b. At the end of the protocol the receiver should learn the string of his choice (i.e., γ_b) but learn nothing about the other string. The sender, on the other hand, should learn nothing about the receiver's choice b.

Oblivious transfer, first introduced by Rabin [Rab81], is a central primitive in modern cryptography. It serves as the basis of a wide range of cryptographic tasks. Most notably, any secure multi-party computation can be based on a secure oblivious transfer protocol [Y86, GMW87, Ki88]. Oblivious transfer has been studied in several variants, all of which have been shown to be equivalent. The variant considered in this paper is the one by Even, Goldreich and Lempel [EGL85] (a.k.a. 1-out-of-2 oblivious transfer), shown to be equivalent to Rabin's original definition by Crépeau [Cr87].

The study of oblivious transfer has been motivated by both theoretical and practical considerations. On the theoretical side, much work has been devoted to the understanding of the hardness assumptions required to guarantee oblivious transfer. We note that known constructions for oblivious transfer are based on relatively strong computational assumptions – either specific assumptions such as factoring or Diffie Hellman (cf. [Rab81, BM89, NP01, AIR01]) or generic assumptions such as the existence of enhanced trapdoor permutations (cf. [EGL85, Go04, Hai04]). Unfortunately, oblivious transfer cannot be reduced in a black box manner to presumably weaker primitives such as one-way functions [IR89]. On the practical side, research has been motivated by the fact that oblivious transfer is considered to be the main bottleneck with respect to the amount of computation required by secure multi-party protocols. This makes the construction of efficient protocols for oblivious transfer a well-motivated task.

In particular, constructing round-efficient oblivious transfer protocols is an important task. Indeed, [NP01] (in Protocol 4.1) and [AIR01] independently constructed a *two-message* (1-round) oblivious transfer protocol based on the DDH Assumption (with weaker security guarantees than the simulation based security). Their work was the starting point of our work.

1.2 Smooth Projective Hashing

Smooth projective hashing was introduced by Cramer and Shoup [CS02]. Informally, a *projective hash family* is a family of keyed hash functions with two types of keys: the primary *hashing key* that can be used to compute the hash function on every point in its domain, and a *projective key* that can only be used to compute the hash function on a "special subset" of its domain. Moreover, to efficiently compute the hash value using the projective key one also needs a "witness" for membership in the special subset. (The domain is typically denoted by X and the special subset is typically denoted by L.) A projective hash family is *smooth* if the projective key gives (almost) no information about the value of the hash function on points outside the special subset. An important property that is used in all the applications of such families is that it is hard to distinguish members of the special subset from non-members. This is called the *hard subset membership* property.

1.3 Oblivious transfer from smooth projective hashing

We present a methodology for constructing a two-message oblivious transfer protocol from any (variant of a) smooth projective hash family. In particular, the protocols of [NP01, AIR01] can be viewed as a special case of this methodology. Moreover, we show that this methodology gives rise to two new oblivious transfer protocols; one based on the Quadratic Residuosity Assumption, and the other based on the N'th Residuosity Assumption.

Our protocols, similarly to the protocols of [NP01, AIR01], are not known to be secure according to the traditional simulation based definition. Yet, they have the advantage of providing a certain level of security even against malicious adversaries without having to compromise on efficiency (see Section 3 for further discussion on the guaranteed level of security).

The basic idea. Given a smooth projective hash family with the hard subset membership property, consider the following two-message oblivious transfer protocol. Recall that the sender S takes as input a pair of strings γ_0, γ_1, and the receiver R takes as input a choice bit b.

$R \to S$: Generate the hashing parameters Λ (that define the domain X and the special subset L). Choose a random triple (x_0, x_1, w_b) where $x_b \in_R L$, w_b is a "witness" for membership of $x_b \in L$, and $x_{1-b} \in_R X \setminus L$. Send (Λ, x_0, x_1).

$S \to R$: Choose independently at random two primary hashing keys k_0, k_1 together with their corresponding projective keys pk_0, pk_1. Send pk_0 and pk_1 along with $y_0 = \gamma_0 \oplus H_{k_0}(x_0)$ and $y_1 = \gamma_1 \oplus H_{k_1}(x_1)$.

R: Retrieve γ_b by computing $y_b \oplus H_{k_b}(x_b)$, using the witness w_b and the projective key pk_b.

The security of the receiver is implied by the hardness of the subset membership problem on X, since guessing the value of b implies distinguishing between a random member and a random non-member. The security of the sender is implied by the smoothness property of the hash family. Specifically, given a random projective key pk and any element in $x \in X \setminus L$, the value $H_k(x)$ is statistically indistinguishable from random. Thus, the message y_{1-b} gives no information about γ_{1-b} (since $x_{1-b} \in X \setminus L$). Note that the functionality of the protocol is implied by the projection property.

Malicious receivers. The above protocol works for an honest-but-curious receiver, but the security of the sender is no longer guaranteed when considering malicious receivers. The reason is that there is no guarantee that the receiver will choose $x_{1-b} \in X \setminus L$. A malicious receiver might choose $x_0, x_1 \in L$ and learn both γ_0 and γ_1. To overcome this problem, we extend the notion of smoothness so that it is possible to verify that the function is smooth on at least one of x_0, x_1 (and in particular at least one of them in not in L). Note that this must hold even if the hashing parameters Λ are maliciously chosen by the receiver.

Implementing this extended notion in the context of the DDH assumption is straightforward [NP01, AIR01]. Loosely speaking, in this case the hashing parameters consist of a prime p and two elements $g_0, g_1 \in \mathbb{Z}_p^*$ of prime order $q|(p-1)$. The hashing domain is $X \stackrel{\text{def}}{=} \langle g_0 \rangle \times \langle g_1 \rangle = \{(g_0^{r_0}, g_1^{r_1}) : r_0, r_1 \in \mathbb{Z}_q\}$, the special subset is $L \stackrel{\text{def}}{=} \{(g_0^r, g_1^r) : r \in \mathbb{Z}_q\}$, and the witness is the exponent r. To enable the sender to verify that two elements x_0, x_1 are not both in L, we instruct the receiver to generate x_0, x_1 by choosing at random two distinct elements $r_0, r_1 \in \mathbb{Z}_q$, setting $x_b = (g_0^{r_0}, g_1^{r_0})$, $w_b = r_0$, and $x_{1-b} = (g_0^{r_0}, g_1^{r_1})$. Notice that x_b is uniformly distributed in L, x_{1-b} is uniformly distributed in $X \setminus L$, and the sender can easily verify that at least one of x_0, x_1 is not in L by merely checking that they agree on their first coordinate and differ on their second coordinate.

1.4 Factoring-based smooth projective hashing

Implementing smooth projective hashing with the extra verifiability property in the context of the Quadratic Residuosity Assumption and the N'th Residuosity Assumption is not as easy. This part contains the bulk of technical difficulties of this work.

On top of providing the additional verifiability property, we were also able to somewhat relax the underlying assumptions that were used in prior work. The factoring-based constructions of smooth projective hashing, in the work of Cramer and Shoup (as well as all subsequent works), were only proven secure for the special case where the RSA-composite in use is a product of safe primes. Namely, they used $N = pq$ where p, q are distinct odd primes such that $p' = (p-1)/2$ and $q' = (q-1)/2$ are also odd primes.[2] This restriction was explained by "technical reasons," but we observe that it is not needed: in fact our factoring-based constructions can be proven secure also for "generic" RSA-composites (i.e., for any $N = pq$ where p, q are two odd primes of the same size).

Moreover, using the same tools we can eliminate the need for safe primes in the CCA encryption schemes that are based on smooth projective hashing (i.e., they too can be implemented without safe primes). Indeed, we observe that encryption schemes in the literature that are based on the N'th Residuosity or Quadratic Residuosity Assumptions (cf. [CS02, GL03, CS03]) remain secure even when the underlying RSA-composite is not chosen as a product of safe primes. In the appendix we explain how the proofs for the existing schemes can be modified to prove this stronger result, and exemplify it in detail for the proof of Camenisch and Shoup from [CS03]. (We did not check whether the same applies also to the password protocols in [GL03] and [CHK$^+$05] or the proofs of correct encryption and decryption from [CS03].)

Eliminating the safe primes. We describe the idea that allows us to eliminate the need for safe primes in the construction of smooth projective hashing based on the N'th Residuosity Assumption. (For the construction based on the Quadratic Residuosity Assumption, the idea is similar but even simpler.)

[2] p' and q' are also called Sophie Germain primes.

The only place where prior work really used the safe-prime condition is in proving the hard subset membership property. Namely, one needs to prove that it is hard to distinguish the special subset of the hashing scheme from the entire domain. In the construction based on the N'th Residuosity Assumption, the hashing parameters are the RSA-composite N and a random N'th residue g modulo N^2, the hashing domain is $Z^*_{N^2}$ and the special subset is the subgroup $\langle g \rangle$ that is generated by g. The safe-prime condition is used to argue that the randomly chosen N'th residue g generates the subgroup of all the N'th residues with high probability.[3] Thus, the N'th Residuosity Assumption immediately implies that it is hard to distinguish the special subset from the entire domain.

When N is just a plain old RSA-composite, we lose the property that a random N'th residue generates the whole subgroup of N'th residues (this subgroup is typically not cyclic). We thus need to work slightly harder to prove the hard subset membership property. In Lemma 2 we show that the N'th Residuosity Assumption implies that it is hard to distinguish the subgroup $L \stackrel{\text{def}}{=} \langle g \rangle$ from the subgroup $X \stackrel{\text{def}}{=} \langle g \rangle \cdot \langle 1 + N \rangle$ of all the elements that can be generated as $x \leftarrow g^r(1+N)^s \bmod N^2$. We therefore get a smooth projective hashing scheme with domain X and special subset L, and we have the hard subset membership property.

One can see, however, that this construction is still missing one aspect that was present in prior notions. Namely, there could be elements of $Z^*_{N^2}$ that are not in the hashing domain X. Moreover, given the parameters N, g and an element $x \in Z^*_{N^2}$, there does not seem to be an easy way of deciding whether or not $x \in X$. This does not pose any problem for the application to oblivious transfer, but it potentially poses some problems in the case of chosen-ciphertext-secure encryption. Luckily, solving this issue in the context of encryption turns out to be quite straightforward, as we show in the appendix.

2 Notations

For any positive integer X we denote the set $\{0, 1, \ldots, X - 1\}$ by either $[X]$ or Z_X. We always denote the security parameter by n. A function $\nu : \mathbb{N} \to [0, 1]$ is said to be negligible if for every polynomial $p(\cdot)$ and for every large enough n, $\nu(n) < 1/p(n)$. For an algorithm A, $y \leftarrow A(x)$ denotes running A on input x and assigning the result to y. If A is randomized, then y is a random variable. We denote by $x \in_R S$ the action of uniformly choosing an element from the set S.

For any two random variables X, Y, we say that X and Y are ϵ-close, denoted $X \stackrel{\epsilon}{\approx} Y$, if $Dist(X, Y) \leq \epsilon$, where $Dist(X, Y)$ denotes the statistical difference between X and Y.[4] We say that the ensembles $X = \{X_n\}_{n \in \mathbb{N}}$ and $Y = \{Y_n\}_{n \in \mathbb{N}}$ are *statistically indistinguishable*, denoted $X \stackrel{\text{s}}{\equiv} Y$, if there exists a negligible function $\epsilon(\cdot)$ such that for every $n \in \mathbb{N}$, the random variables X_n and Y_n are $\epsilon(n)$-close.

We say that the ensembles $X = \{X_n\}_{n \in \mathbb{N}}$ and $Y = \{Y_n\}_{n \in \mathbb{N}}$ are *computationally indistinguishable*, denoted $X \stackrel{\text{c}}{\equiv} Y$, if for every non-uniform probabilistic polynomial-time (PPT) distinguisher D there exists a negligible function $\epsilon(\cdot)$ such that for every $n \in \mathbb{N}$,

$$|\Pr[D(X_n) = 1] - \Pr[D(Y_n) = 1]| < \epsilon(n)$$

[3] Making this argument requires considering only elements with Jacobi symbol $+1$, but this technicality is irrelevant for our current discussion.

[4] Recall that $Dist(X, Y) \triangleq \frac{1}{2} \sum_{s \in S} |\Pr[X = s] - \Pr[Y = s]|$, or equivalently, $Dist(X, Y) \triangleq \max_{S' \subset S} |\Pr[X \in S'] - \Pr[Y \in S']|$, where S is any set that contains the support of both X and Y.

For simplicity, throughout this paper we say that two random variables X_n and Y_n are computationally (or statistically) indistinguishable, meaning that the corresponding distribution ensembles $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are computationally (or statistically) indistinguishable.

3 Security of Oblivious Transfer

Our definition of oblivious transfer is similar to the ones considered in previous works on oblivious transfer in the Bounded Storage Model [DHRS04, CCM98], and similar to the definition considered in [NP01] in the context of their DDH-based two-message oblivious transfer protocol. We remark that the definition below is specific for two-message protocols, as we only deal with such protocols in this work.

Definition 1 (Secure implementation of Oblivious Transfer) *A two-message, two-party protocol (S, R) is said to* securely implement oblivious transfer *for ℓ-bit strings (where $\ell : \mathbb{N} \to \mathbb{N}$) if it is a protocol in which both the sender and the receiver are probabilistic polynomial-time machines that get as input a security parameter n in unary representation. Moreover, the sender gets as input two strings $\gamma_0, \gamma_1 \in \{0,1\}^{\ell(n)}$, the receiver gets as input a choice bit $b \in \{0,1\}$, and the following conditions are satisfied:*

- Functionality: *If the sender and the receiver follow the protocol then for any security parameter n, any two input strings $\gamma_0, \gamma_1 \in \{0,1\}^{\ell(n)}$, and any bit b, the receiver outputs γ_b whereas the sender outputs nothing.*[5]
- Receiver's security: *The ensembles $\{R(1^n, 0)\}_{n\in\mathbb{N}}$ and $\{R(1^n, 1)\}_{n\in\mathbb{N}}$ are computationally indistinguishable; i.e.,*
$$\{R(1^n, 0)\}_{n\in\mathbb{N}} \overset{c}{\equiv} \{R(1^n, 1)\}_{n\in\mathbb{N}}$$
where $R(1^n, b)$ denotes the message sent by the honest receiver with input $(1^n, b)$.
- Sender's security: *There is a negligible function ν such that for any $n > 0$, any two messages $\gamma_0, \gamma_1 \in \{0,1\}^{\ell(n)}$, and any message $q \in \{0,1\}^*$ (from a possibly cheating, not necessarily polynomial-time receiver), it holds that*
$$Dist(S(1^n, \gamma_0, \gamma_1, q),\ S(1^n, \gamma_0, 0^{\ell(n)}, q)) \leq \nu(n) \quad or \quad Dist(S(1^n, \gamma_0, \gamma_1, q),\ S(1^n, 0^{\ell(n)}, \gamma_1, q)) \leq \nu(n)$$
where $S(1^n, \gamma_0, \gamma_1, q)$ denotes the response of the honest sender with input $(1^n, \gamma_0, \gamma_1)$ when the receiver's first message is q.

Note that Definition 1 (similarly to the definitions in [DHRS04, CCM98, NP01]) departs from the simulation-based definition in that it handles the security of the sender and of the receiver separately. This results in a somewhat weaker security guarantee.

The simulation-based definition compares the "real world," where the parties execute the protocol, to an "ideal world," where no message is exchanged between the two parties; rather, there is an "ideal functionality" (or a trusted party) that takes an input from both parties, computes the output of the Oblivious Transfer functionality on these inputs, and sends the corresponding output to each party. Loosely speaking, the simulation-based definition asserts that for every non-uniform

[5] This condition is also referred to as the completeness condition.

probabilistic polynomial-time adversary $\mathcal{A}$ (controlling either the sender of the receiver) in the "real world" there exists a non-uniform probabilistic polynomial-time simulator $\mathcal{S}$, controlling the same party in the "ideal world," that for any input can simulate the view of the adversary $\mathcal{A}$ in a computationally indistinguishable manner.

We note that Definition 1 does give a simulation-based guarantee in the case that the sender is corrupted. In this case the simulator, who does not know the choice bit of the receiver, can simulate the message of the receiver by setting it to be $R(1^n, 0)$.[6] The fact that this view is indistinguishable from the "real world" view follows from the receiver's security which asserts that $\{R(1^n, 0)\}_{n\in\mathbb{N}} \stackrel{c}{\equiv} \{R(1^n, 1)\}_{n\in\mathbb{N}}$.

On the other hand, Definition 1 does not give a simulation-based guarantee in the case that the receiver is corrupted. The reason is that a malicious receiver is not guaranteed to "know" its input bit b, and therefore the simulator does not know which input bit to feed the ideal functionality in order to obtain the desired output γ_b. However, Definition 1 does guarantee an exponential time simulation of the receiver's view of the interaction (similarly to the definition of [NP01]). Loosely speaking, the simulator extracts the bit b from the malicious receiver (in exponential time), feeds it to the ideal functionality, receives an output γ, and simulates the sender's message in the protocol by setting it to be $S(\gamma, 0^{\ell(n)}, q)$ if $b = 0$ and $S(0^{\ell(n)}, \gamma, q)$ if $b = 1$, where q is the first message sent by the malicious receiver. This implies that the interaction gives no information to an unbounded receiver beyond the value of γ_b. We note that if the receiver is semi-honest (a.k.a, honest but curious), then the simulator works in polynomial time, and thus Definition 1 does guarantee simulation-based security for semi-honest adversaries.

4 Smooth Projective Hash Functions

Our definition of smooth projective hashing differs in some ways from prior definitions [CS02, GL03], mainly in that we add the requirement that it is possible to verify that at least one of two given elements is a "non-member." We also depart from the presentation in previous work and define the notion of smooth projective hashing in terms of the procedures that are used to implement it rather than in terms of languages and sets. (This is merely a presentation issue, but we believe that it makes the presentation clearer.) At the end of this section we briefly discuss the mapping between our presentation and the one used in previous work.

Syntax. A hash family $\mathcal{H}$ is defined by means of the following six polynomial-time algorithms, $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$:

- The *parameter-generator* PG is a randomized algorithm that takes as input the security parameter and outputs some parameters, $\Lambda \leftarrow \mathsf{PG}(1^n)$. We sometimes assume for simplicity that $|\Lambda| = n$ whenever Λ is in the support of $\mathsf{PG}(1^n)$.

- The *instance-sampler* IS is a randomized algorithm that takes as input the parameters Λ and outputs a triple, $(w, x, x') \leftarrow \mathsf{IS}(\Lambda)$. The intent is that x is a member of the special subset, x' is a non-member, and w is a witness for the membership of x in the special subset.

[6] Notice that in this case the simulator does not benefit from the ideal functionality since the sender does not receive any output from the ideal functionality.

- The *instance-testing* algorithm IT tests the parameters Λ and two strings x_0, x_1, namely $\mathsf{IT}(\Lambda, x_0, x_1) \in \{0,1\}$. The intent is to test that at least one of x_0, x_1 is not a member in the special subset.

- The *hash-key generator* HG is a randomized algorithm that takes as input the parameters Λ and outputs two keys (i.e., a primary hashing key and a projective key), $(hk, pk) \leftarrow \mathsf{HG}(\Lambda)$.

- The *primary hash algorithm* Hash takes the parameters Λ and hash key k and an element x and outputs a string $y \leftarrow \mathsf{Hash}(\Lambda, hk, x)$.

- The *secondary (projection) hash algorithm* pHash takes the parameters Λ and projective key pk and two elements (w, x) and outputs a string $y \leftarrow \mathsf{pHash}(\Lambda, pk, w, x)$.

For every string Λ, consider using Λ as the hashing parameters and let G_Λ denote the set of possible hash values with these parameters, namely

$$\begin{aligned} G_\Lambda \stackrel{\text{def}}{=} \; & \{\mathsf{Hash}(\Lambda, hk, x) : (w, x, x') \in \mathsf{support}(\mathsf{IS}(\Lambda)),\ (hk, pk) \in \mathsf{support}(\mathsf{HG}(\Lambda))\} \\ \cup \; & \{\mathsf{Hash}(\Lambda, hk, x') : (w, x, x') \in \mathsf{support}(\mathsf{IS}(\Lambda)),\ (hk, pk) \in \mathsf{support}(\mathsf{HG}(\Lambda))\} \end{aligned}$$

It helps to think of G_Λ as a group where we can efficiently compute the group operation and its inverse. In most of this paper, the group G_Λ will be the set of ℓ-bit strings (with the xor operation) where $\ell = \ell(|\Lambda|)$ for some polynomially-bound function $\ell(\cdot)$.

Definition 2 *Let $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$ and let Λ and x be two strings.*

Smoothness. *Let $\epsilon \geq 0$. We say that $\mathcal{H}$ is* ϵ-smooth *on (Λ, x) if the following two distributions are ϵ-close:*

$$\Big[(pk, \mathsf{Hash}(\Lambda, hk, x))\Big]_{(hk,pk)\leftarrow \mathsf{HG}(\Lambda)} \quad \textit{and} \quad \Big[(pk, y)\Big]_{(hk,pk)\leftarrow \mathsf{HG}(\Lambda),\ y \in_R G_\Lambda}$$

The first distribution is induced by choosing $(hk, pk) \leftarrow \mathsf{HG}(\Lambda)$ and outputting $(pk, \mathsf{Hash}(\Lambda, hk, x))$, and the second is induced by choosing independently $y \in_R G_\Lambda$ and $(hk, pk) \leftarrow \mathsf{HG}(\Lambda)$ and outputting (pk, y).

Projection. *We say that $\mathcal{H}$ is* projective *on (Λ, x) if whenever both (hk, pk) and (hk', pk) (with the same pk) are in the support of $\mathsf{HG}(\Lambda)$, it holds that $\mathsf{Hash}(\Lambda, hk, x) = \mathsf{Hash}(\Lambda, hk', x)$.*

It is easy to verify that ϵ-smoothness on (Λ, x) and projection on (Λ, x) are contradictory requirements (assuming that $\epsilon < 1 - \frac{1}{|G_\Lambda|}$).

Definition 3 (Smooth Projective Hashing) *A family $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$ is a smooth projective hash family if there exists a negligible function $\epsilon : \mathbb{N} \to [0,1]$ such that for every $\Lambda \in \mathsf{support}(\mathsf{PG})$, every $(w, x, x') \in \mathsf{support}(\mathsf{IS}(\Lambda))$, and every $(hk, pk) \in \mathsf{support}(\mathsf{HG}(\Lambda))$, it holds that*

(a) $\mathsf{pHash}(\Lambda, pk, w, x) = \mathsf{Hash}(\Lambda, hk, x)$.

(b) $\mathcal{H}$ is $\epsilon(|\Lambda|)$-smooth on (Λ, x').

(Clearly condition (a) above implies in particular that $\mathcal{H}$ is projective on (Λ, x), since the left-hand-side is independent of hk except via the corresponding pk.)

Definition 4 (Verifiable Smoothness) *A smooth projective hash family* $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$ *is verifiably smooth if in addition to* $(a), (b)$ *it holds that:*

(c) For every $\Lambda \in \mathsf{support}(\mathsf{PG})$ *and every* $(w, x, x') \in \mathsf{support}(\mathsf{IS}(\Lambda))$*, it holds that* $\mathsf{IT}(\Lambda, x, x') = \mathsf{IT}(\Lambda, x', x) = 1$.

(d) For every Λ, x_0, x_1 *such that* $\mathsf{IT}(\Lambda, x_0, x_1) = 1$*, it holds that either* $\mathcal{H}$ *is* $\epsilon(|\Lambda|)$*-smooth on* (Λ, x_0) *or it is* $\epsilon(|\Lambda|)$*-smooth on* (Λ, x_1) *(or both).*

Definition 5 (Hard Subset Membership) *A smooth projective hash family* $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$ *is said to have a* hard subset membership *property if the distribution ensembles* $\{A_n\}_{n\in\mathbb{N}}$, $\{B_n\}_{n\in\mathbb{N}}$ *defined below are computationally indistinguishable:*

Distribution A_n*: Choose at random* $\Lambda \leftarrow \mathsf{PG}(1^n)$ *and* $(w, x, x') \leftarrow \mathsf{IS}(\Lambda)$ *and output* (Λ, x, x').

Distribution B_n*: Choose at random* $\Lambda \leftarrow \mathsf{PG}(1^n)$ *and* $(w, x, x') \leftarrow \mathsf{IS}(\Lambda)$ *and output* (Λ, x', x).

Note that the difference between A_n and B_n is just in the order of x and x'. Note also that this condition is stronger than just requiring indistinguishability between members and non-members in the special subset (since x and x' may be chosen in a dependent manner). This stronger requirement is needed for the Oblivious Transfer application.

4.1 Comments

The definitions above are formulated in a way that is convenient for use in our application of Oblivious Transfer, but may make it harder to see the correspondence to the notions that were defined in previous work [CS02, GL03]. We now briefly discuss this correspondence and provide some other clarifications.

Dealing with "bad inputs." We stress from the outset that many of the notations and definitions above do not depend on the inputs to the various algorithms being chosen "the right way." For example, the set G_Λ is well defined even when Λ is not in the support of the parameter-generation algorithm, and similarly the property of $\mathcal{H}$ being ϵ-smooth on (Λ, x) is well-defined for any two strings Λ and x.

In our application to Oblivious Transfer, we will use the hashing values only for instances that pass the instance-testing procedure, and will use that procedure to weed out nonsensical inputs. In particular, if we have some parameters Λ' that are mal-formed (in a recognizable way) we can have the instance-testing always rejecting them, and then the verifiable-smoothness requirement will be vacuous for such parameters.

Hash domain and the "special subset." Previous works presented the definitions in terms of some (parameter dependent) domain X_Λ for the hash function, and a "special subset" of that domain, which is an NP language $L_\Lambda \subset X_\Lambda$. They also required that it be possible to sample both members of L_Λ and non-members in $X_\Lambda \setminus L_\Lambda$. In our case, the special subset L_Λ is the support of the second element in the output of the instance-sampler, and the non-members $X_\Lambda \setminus L_\Lambda$ is the support of the third element, namely

$$\begin{aligned} L_\Lambda &= \{x : \exists\, w, x' \ s.t.\ (w, x, x') \in \mathsf{support}(\mathsf{IS}(\Lambda))\} \\ X_\Lambda \setminus L_\Lambda &= \{x' : \exists\, w, x \ s.t.\ (w, x, x') \in \mathsf{support}(\mathsf{IS}(\Lambda))\}. \end{aligned}$$

Since we require that the hash family is projective on members of the special subset and smooth on non-members, it follows that these two sets are indeed disjoint (assuming that $|G_\Lambda| > 1$).

The sampleable distributions on these sets are the ones induced by the instance-sampler. We comment that for our application we need to choose these elements in a dependent manner, since we need to sample pairs (x, x') that the instance-testing procedure accepts.

Projection and smoothness. The projection definition that we use is the usual one: for $x \in L_\Lambda$ (i.e., x that was output as the second element in the output of $\mathsf{IS}(\Lambda)$), the value of $\mathsf{Hash}(\Lambda, hk, x)$ is determined by the corresponding projective key pk, and moreover it can be efficiently computed given a "witness" w using the algorithm $\mathsf{pHash}(\Lambda, pk, w, x)$. (This implies in particular that if there are a few "witnesses" for the same x, the value of $\mathsf{pHash}(\Lambda, pk, w, x)$ is the same for all of them.)

Our smoothness definition is the per-instance definition of Gennaro-Lindell rather than "random instance" definition of Cramer-Shoup. That is, we need $\mathcal{H}$ to be smooth *for every* non-member x rather than just for a random non-member.

The group G_Λ. Although it is always possible to use $G_\Lambda = \{0,1\}^\ell$ (see discussion after Definition 6 in Section 6) one can sometime gain efficiency by working with other groups. (For example, for constructions based on DDH it is more natural to work with the underlying DDH group.) The properties that we need from G_Λ for our construction to work are the following:

- G_Λ should be a quasigroup[7] where the operation and "its inverse" can be efficiently computed. Namely, there should be polynomial-time algorithms $\mathsf{op}(\Lambda, \cdot, \cdot)$ and $\mathsf{inv}(\Lambda, \cdot, \cdot)$ such that op computes the quasigroup operation in G_Λ and $\mathsf{inv}(\Lambda, \mathsf{op}(\Lambda, x, y), y) = x$ for all $x, y \in G_\Lambda$.
- It should be possible to encode and decode the sender's inputs as elements in G_Λ. If the sender's inputs are ℓ-bit strings, there should be polynomial-time algorithms $\mathsf{Encode}, \mathsf{Decode}$ so that for all $s \in \{0,1\}^\ell$ it holds that $\mathsf{Encode}(\Lambda, s) \in G_\Lambda$ and $\mathsf{Decode}(\Lambda, \mathsf{Encode}(\Lambda, s)) = s$.

Some redundancies. One can observe that the definitions above are somewhat redundant. For example, it is not hard to see that conditions (a), (c) and (d) of Definitions 3 and 4 together imply also condition (b) (assuming that $|G_\Lambda| > 1$). Also if $\mathcal{H}$ has the hard subset membership property then requiring $\mathsf{IT}(\Lambda, x, x') = 1$ in condition (c) of Definition 4 implies that also $\mathsf{IT}(\Lambda, x', x) = 1$ (except perhaps with a negligible probability).

5 Constructing 2-Message OT Protocols

We now show how to construct a two-message Oblivious Transfer protocol from smooth projective hash functions (defined in Section 4).

[7] $(G, \cdot)$ is a quasigroup if for all $a, b \in G$ there exist unique $x, y \in G$ such that $a \cdot x = y \cdot a = b$.

Let $\ell : \mathbb{N} \to \mathbb{N}$ be a (polynomially-bounded, efficiently computable) function, let $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$ be a verifiably-smooth projective hash family with the hard subset membership property (cf. Definitions 3-5), and assume for simplicity that for every setting of the parameters $\Lambda \in \{0,1\}^*$ it holds that $G_\Lambda = \{0,1\}^{\ell(|\Lambda|)}$.[8] (At the end of this section we briefly discuss the (straightforward) modifications that are needed to deal with other domains.)

Let n be the security parameter. Let (γ_0, γ_1) be the input of the sender, where γ_0 and γ_1 are $\ell(n)$-bit strings, and let $b \in \{0,1\}$ be the input of the receiver.

$R \to S$: The receiver chooses the hashing parameters $\Lambda \leftarrow \mathsf{PG}(1^n)$. He then samples random instances $(w, x, x') \leftarrow \mathsf{IS}(\Lambda)$, sets $x_b := x$ and $x_{1-b} := x'$, and sends (Λ, x_0, x_1) to the sender.

$S \to R$: The sender invokes the testing algorithm $\mathsf{IT}(\Lambda, x_0, x_1)$ (to verify that the hashing is smooth on at least one of x_0, x_1). If the test fails then the sender aborts.

Otherwise the sender runs the hash-key generation algorithm twice independently to get $(hk_0, pk_0) \leftarrow \mathsf{HG}(\Lambda)$ and $(hk_1, pk_1) \leftarrow \mathsf{HG}(\Lambda)$, sets $y_0 \leftarrow \gamma_0 \oplus \mathsf{Hash}(\Lambda, hk_0, x_0)$ and $y_1 \leftarrow \gamma_1 \oplus \mathsf{Hash}(\Lambda, hk_1, x_1)$, and sends (pk_0, pk_1, y_0, y_1) to the receiver.

R: The receiver retrieves γ_b by computing $\gamma_b \leftarrow y_b \oplus \mathsf{pHash}(\Lambda, pk_b, w, x)$.

We next prove that the above protocol is secure according to Definition 1. The functionality follows from the fact that $\mathcal{H}$ is projective, which means that the value $\mathsf{Hash}(\Lambda, hk, x_b)$ that the sender computes is equal to the value $\mathsf{pHash}(\Lambda, pk_b, w, x)$ that the receiver computes. The receiver's security follows from the hard subset membership property, which means that it is hard to distinguish between the pairs (x, x') and (x', x). The sender's security follows from verifiable smoothness, which means that for at least one of $b \in \{0,1\}$ the value of $\mathsf{Hash}(\Lambda, hk_b, x_b)$ is (almost) random in $G_\Lambda = \{0,1\}^\ell$, even given the projective key pk_b.

Theorem 1 *The above 2-message OT protocol is secure according to Definition 1, assuming that $\mathcal{H}$ is a verifiably-smooth projective hash family that has the hard subset membership property.*

Proof The functionality trivially follows from $\mathcal{H}$ being projective. Similarly, the receiver's security trivially follows from $\mathcal{H}$ having the hard subset membership property, since $\{R(1^n, 0)\}_{n\in\mathbb{N}} = \{A_n\}_{n\in\mathbb{N}}$ and $\{R(1^n, 1)\}_{n\in\mathbb{N}} = \{B_n\}_{n\in\mathbb{N}}$. Hence a probabilistic polynomial-time sender $\hat{S}$ that can predict with non-negligible advantage the choice bit b when interacting with $R(1^n, b)$ (on infinitely many auxiliary inputs $\{z_n\}_{n\in\mathbb{N}}$ with $|z_n| \le \mathsf{poly(n)}$) is by definition a distinguisher between $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ from Definition 5.

It is left to prove the sender's security. Fix $n \in \mathbb{N}$ and $\gamma_0, \gamma_1 \in \{0,1\}^{\ell(n)}$. Let X be the first message sent by the receiver, and parse $X = (\Lambda, x_0, x_1)$. If X is rejected by the testing algorithm, i.e. $\mathsf{IT}(\Lambda, x_0, x_1) = 0$, then the sender aborts regardless of its input (so the three random variables $S(1^n, \gamma_0, \gamma_1, X)$, $S(1^n, \gamma_0, 0^{\ell(n)}, X)$, and $S(1^n, 0^{\ell(n)}, \gamma_1, X)$ are identical). If $\mathsf{IT}(\Lambda, x_0, x_1) = 1$ then by verifiable smoothness we know that either $\mathcal{H}$ is ϵ-smooth on (Λ, x_0) or it is ϵ-smooth on (Λ, x_1),

[8] Recall that mal-formed Λ's can be handled using the instance-testing algorithm. See discussion in Section 4.1.

for some negligible ϵ. In the latter case we have

$$\begin{aligned}
S(1^n, \gamma_0, \gamma_1, X) &= (pk_0, pk_1, \gamma_0 \oplus \mathsf{Hash}(\Lambda, hk_0, x_0),\ \gamma_1 \oplus \mathsf{Hash}(\Lambda, hk_1, x_1)) \\
&\overset{\epsilon}{\approx} (pk_0, pk_1, \gamma_0 \oplus \mathsf{Hash}(\Lambda, hk_0, x_0),\ \gamma_1 \oplus y)_{y \in_R \{0,1\}^{\ell(n)}} \\
&= (pk_0, pk_1, \gamma_0 \oplus \mathsf{Hash}(\Lambda, hk_0, x_0),\ 0^{\ell(n)} \oplus y)_{y \in_R \{0,1\}^{\ell(n)}} \\
&\overset{\epsilon}{\approx} (pk_0, pk_1, \gamma_0 \oplus \mathsf{Hash}(\Lambda, hk_0, x_0),\ 0^{\ell(n)} \oplus \mathsf{Hash}(\Lambda, hk_1, x_1)) \\
&= S(1^n, \gamma_0, 0^{\ell(n)}, X)
\end{aligned}$$

and in the former case we similarly get $S(1^n, \gamma_0, \gamma_1, X) \overset{2\epsilon}{\approx} S(1^n, 0^{\ell(n)}, \gamma_1, X)$. This concludes the proof, since $\epsilon = \epsilon(n)$ is negligible in n. ■

5.1 Working with different G_Λ

In the protocol above we assumed that for all Λ the group G_Λ is $\{0,1\}^{\ell(|\Lambda|)}$. To work with other (quasi) groups we need the additional properties that were discussed at the end of Section 4.1. The protocol then needs to be modified so that instead of just computing $y_b \leftarrow \gamma_b \oplus \mathsf{Hash}(\Lambda, hk_b, x_b)$, the sender encodes γ_b as an element $\Gamma_b \in G_\Lambda$ and then sends to the receiver $Y_b \leftarrow \Gamma_b \cdot \mathsf{Hash}(\Lambda, hk_b, x_b)$ with '$\cdot$' being the (quasi) group operation. Similarly, the receiver computes the "inverse operation" $\Gamma_b \leftarrow \mathsf{inv}(\Lambda, Y_b, \mathsf{pHash}(\Lambda, pk_b, w, x))$ and decode Γ_b to get γ_b.

The security of this protocol is proven similarly to Theorem 1. The arguments for functionality and receiver security are exactly as before. As for the sender security, we now replace $\gamma \oplus y$ for $y \in_R \{0,1\}^{\ell(n)}$ by $\Gamma \cdot Y$ for $y \in_R G_\Lambda$. Since G_Λ is a quasigroup, this last distribution is the uniform distribution over G_Λ, regardless of what Γ is, and the proof follows.

6 Constructing Smooth Projective Hash Families

We next present two constructions of verifiably-smooth projective hash families with the hard subset membership property. In one construction the hard subset membership property is based on the Quadratic Residuosity Assumption, and in the other the hard subset membership property is based on the N'th Residuosity Assumption. A key vehicle in both constructions is the notion of a verifiably-ϵ-universal projective hash family.

Definition 6 (Universal Hashing) *Let $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$, let Λ and x be two strings, and let $\epsilon > 0$. We say that $\mathcal{H}$ is ϵ-*universal *on (Λ, x) if for any $y_0 \in G_\Lambda$ and any pk_0 it holds that*

$$\Pr_{(hk,pk)}[pk = pk_0 \text{ and } \mathsf{Hash}(\Lambda, hk, x) = y_0] \quad \leq \quad \epsilon \cdot \Pr_{(hk,pk)}[pk = pk_0]$$

where the probability is taken over a random choice $(hk, pk) \leftarrow \mathsf{HG}(\Lambda)$.

Definition 7 (Verifiably-ϵ-universal Projective Hashing) *Let $\epsilon(\cdot)$ be a function. A family $\mathcal{H} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$ is an ϵ-universal projective hash family if for every Λ in the support of PG, every (w, x, x') in the support of $\mathsf{IS}(\Lambda)$, and every (hk, pk) in the support of $\mathsf{HG}(\Lambda)$, it holds that* *(a)* $\mathsf{pHash}(\Lambda, pk, w, x) = \mathsf{Hash}(\Lambda, hk, x)$.

(b′) $\mathcal{H}$ *is* $\epsilon(|\Lambda|)$*-universal on* (Λ, x').

We say that $\mathcal{H}$ *is* *verifiably-*ϵ*-universal* *projective hash family if in addition:*

(c) For every Λ *in the support of* PG *and every* (w, x, x') *in the support of* $\mathsf{IS}(\Lambda)$*, it holds that* $\mathsf{IT}(\Lambda, x, x') = \mathsf{IT}(\Lambda, x', x) = 1$.

(d′) For every Λ, x_0, x_1 *such that* $\mathsf{IT}(\Lambda, x_0, x_1) = 1$*, it holds that either* $\mathcal{H}$ *is* $\epsilon(|\Lambda|)$*-universal on* (Λ, x_0) *or it is* $\epsilon(|\Lambda|)$*-universal on* (Λ, x_1) *(or both).*

Note that conditions (a) and (c) are identical to the ones in Definitions 3 and 4, and conditions (b′) and (d′) only differ in that we replaced ϵ-smooth by ϵ-universal (and ϵ need not be negligible).

Cramer and Shoup have shown in [CS02] how to transform an ϵ-universal projective hash family into a smooth projective hash family (for any $\epsilon < 1$), and the same transformation also works for transforming verifiably-ϵ-universal projective families into verifiably-smooth projective families. In a nutshell, one first reduces ϵ to ϵ^t by choosing t independent hashing key-pairs and hashing t times the same element. One then uses a "strong randomness extractor" [NZ96] to extract a nearly uniform bit string from the t hash values. The new hash algorithms thus use t keys of the original algorithms and also the seed s for the extractor, setting

$$\begin{aligned}\mathsf{Hash}'(\Lambda, (s, hk_1, \ldots, hk_t), x) &= \mathsf{Extract}(s; \mathsf{Hash}(\Lambda, hk_1, x), \ldots, \mathsf{Hash}(\Lambda, hk_t, x)) \\ \mathsf{pHash}'(\Lambda, (s, pk_1, \ldots, pk_t), w, x) &= \mathsf{Extract}(s; \mathsf{pHash}(\Lambda, pk_1, w, x), \ldots, \mathsf{pHash}(\Lambda, pk_t, w, x))\end{aligned}$$

To extract ℓ bits, it is sufficient to choose t so that $t \cdot \log(1/\epsilon) \geq \ell + \omega(\log(|\Lambda|))$. We comment that the resulting construction always has $G'_\Lambda = \{0,1\}^\ell$, regardless of the group G_Λ of the original construction. Further details are omitted and the reader is referred to [CS02]. We conclude that to prove existence of verifiably-smooth projective hash families it suffices to construct verifiably-ϵ-universal projective hash families. In the remaining of this paper we present two such constructions, one based on the N'th Residuosity Assumption and the other based on the Quadratic Residuosity Assumption. Both schemes are obtained by modifying the universal projective schemes of Cramer and Shoup to add the verifiable-universality property (and also to improve some parameters).

6.1 A construction based on the N'th Residuosity Assumption

Let p, q be two distinct odd primes, let $N = pq$ and let R_N be the subgroup consisting of all N'th powers of elements in the multiplicative group $Z^*_{N^2}$. The N'th Residuosity Assumption, originally introduced by Paillier [Pa99], asserts (informally) that given only N, it is hard to distinguish random elements of $Z^*_{N^2}$ from random elements of R_N.

The Cramer-Shoup Scheme. Cramer and Shoup constructed in [CS02] an ϵ-universal projective hash family from the N'th Residuosity Assumption (in the special case where N is a product of two safe primes). Omitting some details, the hash parameters Λ are the modulus $N = pq$ and an N'th residue $g \in R_N$ (with Jacobi symbol $+1$), the hashing key is a random integer $k \in_R \{1, 2, \ldots, \lfloor N^2/2 \rfloor\}$, the projective key is $pk = g^k \bmod N^2$, and the hash is computed as $\mathsf{Hash}(N, g, k, x) = x^k \bmod N^2$. The "special subset" is $L_{N,g} = \{g^w : w < N/2\}$, the exponent w is a "witness" for $x \in L_{N,g}$, and the "non-members" are those elements $x \in Z^*_{N^2}$ (with Jacobi symbol +1) whose order is divisible by p or q.[9] Given the witness w for the element x and the projective key pk, one can compute the hash value as

$$\mathsf{pHash}(N, g, pk, w, x) = pk^w = g^{kw} = x^k \mod N^2$$

Cramer and Shoup proved that when N is a product of two safe primes, this scheme is an ϵ-universal projective family with $\epsilon \approx \max\{1/p,\ 1/q\}$). Moreover, for that case they reduced distinguishing members from non-members to the N'th Residuosity Assumption. In their proofs they make strong use of the fact that when N is a product of two safe primes, the N'th residues are exactly those elements of $Z^*_{N^2}$ whose order is co-prime to N, and moreover the subgroup of N'th residues with Jacobi symbol +1 is cyclic.

Our Modifications. In our case we must also consider a maliciously chosen modulus N that is not necessarily a product of two primes. To get verifiable smoothness (or verifiable universality) we need a way of checking that (a) the order of the element g is co-prime with N, and (b) the order of at least one of the two given elements x_0, x_1 is *not co-prime* with N. For the former, since we do not know how to test that g belongs to the subgroup of elements whose order is co-prime with N, we instead force it into that subgroup by raising it to the power of $N^{\lceil 2\log N\rceil}$. Namely, instead of using the element g itself we use the element $g^{N^{\lceil 2\log N\rceil}} \bmod N^2$.

The latter verification can in principle be done simply by checking that the order of x_1/x_0 in $Z^*_{N^2}$ is divisible by N (and greater than one), which implies that if one of them has order that is co-prime with N then the other one necessarily does not. In our scheme, however, we use a slightly more elaborate test that yields better universality bound ϵ. Specifically, we test that x_1/x_0 is of the form $(1 + vN)$ where v and N are co-primes. Other modifications that we made to the Cramer-Shoup scheme are (a) we eliminate the need of safe primes and (b) we do not restrict our attention to elements with Jacobi symbol +1.

6.1.1 Detailed Construction

Using our notations from Section 4, we now describe the six algorithms that define the hash family $\mathcal{H}_{NR} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$.

Parameter-generator $\mathsf{PG}(1^n)$**.** Choose two random n-bit prime numbers p, q (with $p < q < 2p$). Set $N \leftarrow pq$, choose an element $g' \in_R Z^*_{N^2}$, set $g \leftarrow (g')^N \bmod N^2$, and output $\Lambda = (N, g)$.

Instance-sampler $\mathsf{IS}(N, g)$**.** Choose $v, w \in_R Z^*_N$, compute $T \leftarrow N^{\lceil 2\log N\rceil}$, $x \leftarrow g^{Tw} \bmod N^2$ and $x' \leftarrow x \cdot (1 + vN) \bmod N^2$. Output (w, x, x').

Instance-testing algorithm $\mathsf{IT}(N, g, x, x')$**.** Check that $N > 2^{2n}$ and that $g, x \in Z^*_{N^2}$. Then set $d \leftarrow x'/x \bmod N^2$, verify that $(d - 1)$ is divisible by N (over the integers) and set $v \leftarrow (d - 1)/N$. Finally, verify that v and N are co-primes. Output '1' if all the tests pass and '0' otherwise.

Hash-key generator $\mathsf{HG}(N, g)$**.** Choose $k \in_R Z_{N^2}$, set $T \leftarrow N^{\lceil 2\log N\rceil}$ and $pk \leftarrow (g^T)^k \bmod N^2$. Output (k, pk).

Primary hashing algorithm $\mathsf{Hash}(N, g, k, x)$**.** Output $x^k \bmod N^2$.

Projective hash algorithm $\mathsf{pHash}(N, g, pk, w, x)$**.** Output $pk^w \bmod N^2$.

[9]Note that Paillier's assumption implies that it is hard to distinguish a random element in $Z^*_{N^2}$ with Jacobi symbol +1 from a random N'th residue with Jacobi symbol +1.

6.1.2 Proof of Security

Before proving that the construction above satisfies the properties that we need, we recall some useful facts about the structure of $Z_{N^2}^*$ and restate the N'th Residuosity Assumption.

Fact 6.1 *Let N be a positive integer and let $Z_{N^2}^*$ be the multiplicative group of integers modulo N^2. Then the set $G_N \stackrel{\text{def}}{=} \{1 + vN : v \in Z_N\}$ is a subgroup of $Z_{N^2}^*$ which is homomorphic to the additive group Z_N. In particular, the order of $(1 + vN)$ in $Z_{N^2}^*$ equals the order of v in Z_N.*

Assumption 8 (N'th Residuosity [Pa99]) *The following ensembles are computationally indistinguishable*

$$\begin{aligned} & \{(N, x) \;:\; p, q \in_R \mathsf{Primes}(n),\; N \leftarrow pq,\; x \in_R Z_{N^2}^*\}_{n \in \mathbb{N}} \\ \stackrel{\mathsf{c}}{\equiv}\; & \{(N, y) \;:\; p, q \in_R \mathsf{Primes}(n),\; N \leftarrow pq,\; x \in_R Z_{N^2}^*,\; y \leftarrow x^N \bmod N^2\}_{n \in \mathbb{N}} \end{aligned}$$

where $\mathsf{Primes}(n)$ *denotes the set of prime numbers between 2^n and 2^{n+1}.*

Lemma 2 *Under the N'th Residuosity Assumption, the construction of $\mathcal{H}_{NR}$ from Section 6.1.1 has the hard subset membership property.*

Proof We need to prove indistinguishability between the ensembles

$$A_n = \langle N, g, x, x' \rangle_n \text{ and } B_n = \langle N, g, x', x \rangle$$

where both are taken over choosing

$$\begin{aligned} & p, q \in_R \mathsf{Primes}(n), \quad N \leftarrow pq, \quad T \leftarrow N^{\lceil 2 \log N \rceil}, \\ & g' \in_R Z_{N^2}^*, \quad g \leftarrow (g')^N \bmod N^2, \\ & w, v \in_R Z_N^*, \quad x \leftarrow g^{Tw} \bmod N^2, \quad x' \leftarrow x(1 + vN) \bmod N^2 \end{aligned}$$

Assume for the sake of contradiction that there exists a PPT algorithm D that distinguishes between the ensembles A_n and B_n with non-negligible probability. Let

$$p_1(n) \stackrel{\text{def}}{=} \Pr[D(N, g, x, x') = 1], \quad p_2(n) \stackrel{\text{def}}{=} \Pr[D(N, g, x', x) = 1], \quad \text{and } \epsilon(n) \stackrel{\text{def}}{=} |p_1(n) - p_2(n)|$$

Assume without loss of generality that there is an infinite set $S \subseteq \mathbb{N}$ such that for every $n \in S$ it holds that $p_1(n) \geq p_2(n)$, and moreover $\epsilon(n) = p_1(n) - p_2(n) \geq \frac{1}{\mathsf{poly(n)}}$. We now describe a PPT distinguisher D' for the N'th Residuosity Assumption, that for every $n \in S$, has an advantage (close to) $\epsilon(n)/2$, using D as a subroutine. $D'(N, z)$ works as follows:

1. Choose a random bit $b \in_R \{0, 1\}$ and $v, w \in_R Z_N^*$.
2. Set $g \leftarrow z^N \bmod N^2$, $x \leftarrow z^w \bmod N^2$, and $x' \leftarrow x(1 + vN) \bmod N^2$.
3. Set $x_b \leftarrow x$ and $x_{1-b} \leftarrow x'$, run D to get $b' \leftarrow D(N, g, x_0, x_1)$, and output $b' \oplus b$.

We first claim that when the input z is chosen as a random N'th residue modulo N^2 then the distribution on (N, g, x, x') is nearly identical to the distribution in the actual scheme. To see this, first note that since N was chosen as $N \leftarrow pq$ where p, q are primes and $p < q < 2p - 1$, then N and $\varphi(N)$ are co-primes. Next, we make the simplifying assumption that the exponent w is chosen at random in $Z_{\varphi(N)}$ instead of from Z_N^* (in both the reduction and the instant-sampling algorithm of the scheme itself). It is well known that this modification changes the various distributions by only $O(2^{-n})$ (where n is the bit-length of p, q).

Since N and $\varphi(N)$ are co-primes, then multiplication by powers of N is a permutation modulo $\varphi(N)$, which implies also that exponentiation to powers of N is a permutation on R_N. Hence choosing $g \in_R R_N$ (as done in the scheme) induces the same distribution as choosing $z \in_R R_N$ and setting $g \leftarrow z^N \bmod N^2$ (as in the reduction). Moreover, choosing $w \in_R Z_{\varphi(N)}$ and setting $x \leftarrow g^{Tw \bmod \varphi(N)} \bmod N^2$ (as in the scheme) induces the same distribution as choosing $w \in_R Z_{\varphi(N)}$ and setting $x \leftarrow z^w = g^{N^{-1}w \bmod \varphi(N)} \bmod N^2$ (as in the reduction). Finally, computing x' from x is done in exactly the same way in the scheme and in the reduction. We thus get that for every $n \in S$,

$$\begin{aligned}
\Pr_{R_N}[D'(N,z)=1] &\geq \Pr[b=0]\cdot\Pr[D(N,g,x,x')=1] + \Pr[b=1]\cdot\Pr[D(N,g,x',x)=0] - O(2^{-n}) \\
&= \frac{1}{2}\cdot\big(\Pr[D(N,g,x,x')=1] + (1-\Pr[D(N,g,x',x)=1])\big) - O(2^{-n}) \\
&= \frac{1+\epsilon(n)}{2} - O(2^{-n})
\end{aligned}$$

where by $\Pr_{R_N}[\cdot]$ we mean the probability over the choice of N and a random choice of $z \in_R R_N$.

We next show that when z is a random element in Z_N^* then the input distribution of the subroutine D of D' is independent of the bit b. We observe that setting $x' \leftarrow x(1+vN) = z^w(1+vN) \bmod N^2$ induces the same distribution as setting $x' \leftarrow x(1+vwN) = (z(1+vN))^w \bmod N^2$. (Since v, w are uniform and independent in Z_N^*, and thus so are w, vw). We therefore consider this alternate setting of x' in the analysis. Notice that when z is a random element in Z_N^*, the distribution of the random variable $(N, z, z(1+vN))$ is identical to the distribution of the random variable $(N, z(1+vN), z)$. This implies that the distribution of the random variable $(N, z^w, (z(1+vN))^w)$ is identical to the distribution of the random variable $(N, (z(1+vN))^w, z^w)$. Finally, it remains to notice that since $z^N = (z(1+vN))^N \bmod N^2$, it holds that the distribution of the random variable $(N, z^N, z^w, (z(1+vN))^w) = (N, g, x, x')$ is identical to the distribution of the random variable $(N, z^N, (z(1+vN))^w, z^w) = (N, g, x', x)$. This in turn implies that the output bit b' of the subroutine D of D' must be independent of the bit b, and thus that $\Pr[D'(N,z)=1] = \frac{1}{2}$.

Combining the analysis for the two cases, we get that for every $n \in S$,

$$\Pr_{R_N}[D'(N,z)=1] - \Pr_{Z_{N^2}^*}[D'(N,z)=1] \geq \frac{1+\epsilon(n)}{2} - O(2^{-n}) - \frac{1}{2} = \frac{\epsilon(n)}{2} - O(2^{-n})$$

■

The following immediate corollary is not needed for our application to Oblivious Transfer, but it is useful for other purposes (such as the application to CCA secure encryption that is described in the appendix):

Corollary 3 *Under the N'th Residuosity Assumption, the uniform distribution on* $\langle g^N \rangle \subseteq Z^*_{N^2}$ *is indistinguishable from the uniform distribution on* $\langle g^N \rangle \cdot \langle 1+N \rangle \subseteq Z^*_{N^2}$*, given a random RSA-composite* N *and a random N'th residue* $g^N \in Z^*_{N^2}$ *(where* $g \in_R Z^*_{N^2}$*). More precisely, the following two ensembles are computationally indistinguishable:*

$$\langle N, g^N, g^{Nr} \rangle \stackrel{c}{\equiv} \langle N, g^N, g^{Nr}(1+N)^s \rangle$$

where both are taken over choosing

$$p, q \in_R \mathsf{Primes}(n), \quad N \leftarrow pq, \quad g \in_R Z^*_{N^2}, \quad r \in_R Z_{\varphi(N)}, \; s \in_R Z_N$$

Lemma 4 *The construction of* $\mathcal{H}_{NR}$ *from Section 6.1.1 is a verifiable-*ϵ*-universal projective hash family where* $\epsilon(n) \le 2^{-n} + 2^{-2n}$*.*

Proof The projection and the completeness of the scheme (conditions (a) and (c) in Definition 7) are easy to check: For any RSA modulus N and any $g \in R_N$, $w \in Z^*_N$, and $k \in Z_{N^2}$, setting $x = g^{Tw} \bmod N^2$ and $pk = (g^T)^k \bmod N^2$ (where $T = N^{\lceil 2\log N \rceil}$) we get for Condition (a):

$$\mathsf{pHash}(N, g, pk, w, x) = pk^w = (g^{Tk})^w = (g^{Tw})^k = x^k = \mathsf{Hash}(N, g, k, x) \mod N^2$$

As for Condition (c), the instance-sampler sets $x' \leftarrow x(1+vN) \bmod N^2$ for some $v \in Z^*_N$. Condition (c) follows from the fact that $\frac{(x'/x)-1}{N} = v$, $\frac{(x/x')-1}{N} = -v$, and both v and $-v$ are co-prime with N.

The more interesting property is the verifiable-ϵ-universality (property (d')), which we prove next. (Note that properties (a), (c), and (d') together imply also property (b').) Fix any (N, g, x, x') such that $\mathsf{IT}(N, g, x, x') = 1$. Namely, $N > 2^{2n}$, $g, x \in Z^*_{N^2}$, and $x' = x(1+vN) \mod N^2$ for some $v \in Z^*_N$ (which means that also $x' \in Z^*_{N^2}$). We show that either $\mathcal{H}_{NR}$ is ϵ-universal on (N, g, x) or it is ϵ-universal on (N, g, x').

Recall that the hashing key is chosen as $k \in_R Z_{N^2}$, the projective key is then computed as $pk \leftarrow g^{Tk} \bmod N^2$ where $T = N^{\lceil 2\log N \rceil}$, and the hash function is computed as $\mathsf{Hash}(N, g, k, x) = x^k \bmod N^2$. We make the simplifying assumption that the hashing key k is chosen from $Z_{\varphi(N^2)}$ instead of from Z_{N^2}, thus introducing an error of $O(1/\sqrt{N}) = O(2^{-n})$ into the analysis. For the rest of the proof denote $\tau \stackrel{\text{def}}{=} \text{ord}(g^T)$ and observe that τ must be co-prime with N and therefore must divide $\varphi(N)$.

Consider now the following procedure for choosing the hashing key k, that implies the same distribution as choosing $k \in_R Z_{\varphi(N^2)}$: First choose $k_0 \in_R \{0, \ldots, \tau-1\}$, then $k_1 \in_R \{0, 1, \ldots, \frac{\varphi(N^2)}{\tau} - 1\}$, and then set $k \leftarrow k_0 + \tau k_1$. Observe that the projective key pk depends only on the choice of k_0, since $g^{T(k_0+\tau\cdot k_1)} = g^{Tk_0} \mod N^2$. Below we prove, however, that the hash value on at least one of x, x' must depend also on the choice of k_1.

Denote $d \stackrel{\text{def}}{=} (x'/x) = 1 + vN \mod N^2$, and recall from Fact 6.1 that since v is co-prime with N then the order of d in $Z^*_{N^2}$ is exactly N. Let $\alpha \stackrel{\text{def}}{=} GCD(N, \text{ord}(x))$. Thus $\gamma \stackrel{\text{def}}{=} N/\alpha$ divides the order of x'.

We now show that $\mathcal{H}$ is $(\frac{1}{\gamma} + \frac{1}{N})$-universal on (N, g, x'), and a similar argument shows that it is also $(\frac{1}{\alpha} + \frac{1}{N})$-universal on (N, g, x). Observing that $N > 2^{2n}$ and thus either α or γ must be larger than 2^n completes the proof.

Recall the alternate procedure from above for choosing $k \leftarrow k_0 + \tau k_1$, which implies that

$$\mathsf{Hash}(N, g, k, x') = (x')^k = (x')^{k_0} \cdot (x')^{\tau k_1} \mod N^2.$$

Also recall that τ (which is the order of g^T) is co-prime with N. Thus τ is also co-prime with γ since γ divides N. Hence for two different values $k_1 \neq k_1' \mod \gamma$ we have $\tau k_1 \neq \tau k_1' \mod \gamma$ and therefore also $\tau k_1 \neq \tau k_1' \mod \mathrm{ord}(x')$ (since γ also divides $\mathrm{ord}(x')$), which means that $(x')^{\tau k_1} \neq (x')^{\tau k_1'} \mod N^2$.

Next notice that $\tau | \varphi(N)$ implies $\tau \leq \varphi(N)$ and so $\frac{\varphi(N^2)}{\tau} - 1 \geq N - 1 \geq \gamma - 1$. It follows that when choosing $k_1 \in_R \{0, 1, \ldots, \frac{\varphi(N^2)}{\tau} - 1\}$, the random variable ($k_1 \mod \gamma$) can assume all the values between 0 and $\gamma - 1$, so $(x')^{k_1 \tau}$ can assume at least γ different values modulo N^2. Moreover, each value of ($k_1 \mod \gamma$) occurs either $\lfloor \frac{\varphi(N^2)}{\tau} - 1/\gamma \rfloor$ or $\lceil \frac{\varphi(N^2)}{\tau} - 1/\gamma \rceil$ times, and with $\frac{\varphi(N^2)}{\tau} - 1 \geq N - 1$ it means that no value has probability of more than $\frac{1}{\gamma} + \frac{1}{N-1}$. Hence $\mathcal{H}$ is $(\frac{1}{\gamma} + \frac{1}{N-1})$-universal on (N, g, x')

A symmetric argument shows that $\mathcal{H}$ is $(\frac{1}{\alpha} + \frac{1}{N-1})$-universal on (N, g, x). Since $\alpha\gamma = N > 2^{2n}$ then at least one of α, γ must be larger than 2^n, hence $\mathcal{H}$ is $(2^{-n} + 2^{-2n})$-universal on at least one of (N, g, x) or (N, g, x'). ■

6.2 A construction based on the Quadratic Residuosity Assumption

Let p, q be distinct odd primes; Let $N = pq$ and let QR_N be the subgroup consisting of all squares of elements in the multiplicative group Z_N^*. Also, let J_N be the subgroup of Z_N^* consisting of all elements with Jacobi symbol 1. The Quadratic Residuosity Assumption asserts (informally) that given only N, it is hard to distinguish random elements of J_N from random elements of QR_N.

The Cramer-Shoup Scheme. Cramer and Shoup constructed in [CS02] an ϵ-universal projective hash family from the Quadratic Residuosity Assumption (in the special case where N is a product of two safe primes). This construction is very similar to the N'th Residuosity construction, with the group of N residues modulo N^2 replaced with the group of quadratic residues modulo N. Again, omitting some details the hash parameters are the modulus $N = pq$ and a quadratic residue $g \in_R QR_N$, the hashing key is a random integer $k \in_R \{1, 2, \ldots, N/2\}$, the projective key is $pk = g^{2k} \mod N$, the "special subset" is $L_{N,g} = \{g^{2w} : w < N/2\}$, the exponent w is a "witness" for $x \in L_{N,g}$, and the "non-members" are those elements $x \in J_N$ whose order is even. Given the witness w for the element x and the projective key pk, one can compute the hash value as

$$\mathsf{pHash}(N, g, pk, w, x) = pk^w = g^{2kw} = x^k \mod N$$

Cramer and Shoup proved that when N is a product of two safe primes, this scheme is an (1/2)-universal projective family. Moreover, they also proved that when N is a product of two safe primes, distinguishing members from non-members can be reduced to the Quadratic Residuosity Assumption. Similarly to the N'th Residuosity construction, here too the proofs make strong use of the fact that when N is a product of two safe primes, the subgroup of quadratic residues is cyclic and consists of exactly those elements in Z_N^* whose order is odd.

Our Modifications. We modify the Cramer-Shoup scheme to handle moduli that are not product of safe primes, similarly to the way we did for N'th residues. Specifically, to get verifiable universality we (a) force the element g into the group of odd-order elements by raising it to the power of $2^{\lceil \log N \rceil}$; and (b) verify that at least one of the two element x_0, x_1 has even order by checking that $x_1/x_0 = -1 \bmod N$. Just like in the case of N'th residues, here too forcing g into "the special subgroup" eliminates the need to rely on safe primes.[10]

6.2.1 Detailed Construction

Using our notations from Section 4, we now describe the six algorithms that define the hash family $\mathcal{H}_{QR} = (\mathsf{PG}, \mathsf{IS}, \mathsf{IT}, \mathsf{HG}, \mathsf{Hash}, \mathsf{pHash})$.

Parameter-generator $\mathsf{PG}(1^n)$. Choose at random two n-bit prime numbers p, q, with $p < q < 2p - 1$ and $p \equiv q \equiv 3 \mod 4$. Set $N \leftarrow pq$, choose at random an element $g' \in_R Z_N^*$, set $g = (g')^2 \bmod N$, and output $\Lambda = (N, g)$.

Instance-sampler $\mathsf{IS}(N, g)$. Choose at random $w \in_R Z_N$, compute $T \leftarrow 2^{\lceil \log N \rceil}$, $x \leftarrow g^{Tw} \bmod N$ and $x' \leftarrow N - x$. Output (w, x, x').

Instance-testing algorithm $\mathsf{IT}(N, g, x, x')$. Check that $N > 2^{2n}$, $g, x \in Z_N^*$, and $x' = N - x$. Output '1' if all the tests pass and '0' otherwise.

Hash-key generator $\mathsf{HG}(N, g)$. Choose at random $k \in_R Z_N$, set $T \leftarrow 2^{\lceil \log N \rceil}$ and $pk \leftarrow (g^T)^k \bmod N$. Output (k, pk).

Primary hashing algorithm $\mathsf{Hash}(N, g, k, x)$. Output $x^k \bmod N$.

Projective hash algorithm $\mathsf{pHash}(N, g, pk, w, x)$. Output $pk^w \bmod N$.

6.2.2 Proof of Security

We now show that the construction above has a hard subset membership domain under the Quadratic Residuosity Assumption, and that it is a verifiable-ϵ-universal projective hash family with $\epsilon \approx \frac{1}{2}$. We begin by recalling the Quadratic Residuosity Assumption.

Assumption 9 (Quadratic Residuosity) *The following ensembles are computationally indistinguishable*

$$\begin{aligned} & \{(N, x) \; : \; p, q \in_R \mathsf{Primes}(n),\; N = pq,\; x \in_R J_N\}_{n \in \mathbb{N}} \\ \stackrel{c}{\equiv} \; & \{(N, y) \; : \; p, q \in_R \mathsf{Primes}(n),\; N = pq,\; x \in_R Z_N^*,\; y \leftarrow x^2 \bmod N\}_{n \in \mathbb{N}} \end{aligned}$$

where $\mathsf{Primes}(n)$ denotes the set of prime numbers between 2^n and 2^{n+1} and J_N is the subgroup of Z_N^* of elements with Jacobi symbol +1.

Lemma 5 *Under the Quadratic Residuosity Assumption, the construction of $\mathcal{H}$ from Section 6.2.1 has the hard subset membership property.*

Proof We need to prove indistinguishability between the ensembles

$$A_n = \langle N, g, x, x' \rangle_n \text{ and } B_n = \langle N, g, x', x \rangle$$

[10] We note that the trick of forcing g into the odd-order subgroup as a way of eliminating the need for moduli of a special form was used also in [Hal99].

where both are taken over choosing

$$p, q \in_R \mathsf{Primes}(n) \text{ s.t. } p \equiv q \equiv 3 \mod 4, \quad N \leftarrow pq, \quad T \leftarrow 2^{\lceil \log N \rceil},$$
$$g' \in_R Z_N^*, \quad g \leftarrow (g')^2 \bmod N, \quad w \in_R Z_N, \quad x \leftarrow g^{Tw} \bmod N, \quad x' \leftarrow N - x$$

Assume for the sake of contradiction that there exists a PPT algorithm D that distinguishes between the ensembles A_n and B_n with non-negligible probability. Let

$$p_1(n) \stackrel{\text{def}}{=} \Pr[D(N, g, x, x') = 1], \quad p_2(n) \stackrel{\text{def}}{=} \Pr[D(N, g, x', x) = 1], \quad \text{and } \epsilon(n) \stackrel{\text{def}}{=} |p_1(n) - p_2(n)|$$

We begin by observing that when choosing $p, q \in_R \mathsf{Primes}(n)$, there is a non-negligible probability to get $p \equiv q \equiv 3 \mod 4$ (in which case the modulus N is called *a Blum integer*). Hence, if the Quadratic Residuosity Assumption holds, it must also hold for Blum integers. From now on we therefore assume the Quadratic Residuosity Assumption for Blum integers. Next we recall that when N is a Blum integer, then the quadratic residues modulo N are exactly these elements that have odd order in Z_N^*. Moreover, every quadratic residue has exactly four square roots modulo N. If (r_1, r_2, r_3, r_4) are the four square roots of some $x \in QR_N$ then exactly one of them is itself a quadratic residue modulo N (call it r_1), and exactly two have Jacobi symbol $+1$ modulo N, namely r_1 and $N - r_1$.

Assume without loss of generality that there is an infinite set $S \subseteq \mathbb{N}$ such that for every $n \in S$ it holds that $p_1(n) \geq p_2(n)$, and moreover $\epsilon(n) = p_1(n) - p_2(n) \geq \frac{1}{\mathsf{poly(n)}}$. We now describe a PPT distinguisher D' for Quadratic Residuosity modulo Blum integers that for every $n \in S$ has an advantage (close to) $\epsilon(n)$, using D as a subroutine. $D'(N, z)$ works as follows:

1. Choose $w \in_R Z_N$ and set $T \leftarrow 2^{\lceil \log N \rceil}$.
2. Set $g \leftarrow z^2 \bmod N$, $x_0 \leftarrow z \cdot g^{Tw} \bmod N$ and $x_1 \leftarrow N - x$.
3. Run D to get $b \leftarrow D(N, g, x_0, x_1)$, and output b.

As was done in the proof of Lemma 2, below we analyze the reduction under the simplifying assumption that the exponent w is chosen uniformly in $Z_{\varphi(N)}$ rather than in Z_N, both in the scheme and in the reduction. Under this assumption, we show that when z is a quadratic residue the input to D is distributed according to A_n, and when z is a quadratic non-residue the input to D is distributed according to B_n. It thus follows that when z is a random element in QR_N then D' outputs 1 with probability $p_1(n)$ and when z is a random element in J_n then D' outputs 1 with probability $p_2(n)$, so the advantage of D' is $p_1(n) - p_2(n) = \epsilon(n)$ (minus the negligible deviation caused by our simplifying assumption).

In both cases, the modulus N is chosen just like in the scheme $\mathcal{H}_{QR}$ and the element g is a random quadratic residue modulo N. It is left to show that x_0, x_1 in the reduction are distributed like (x, x') in the scheme when z is a quadratic residue and like x', x in the scheme when z is a quadratic non-residue.

Case 1: $z \in QR_N$**.** In this case we know that $\beta \stackrel{\text{def}}{=} \mathrm{ord}_N(z)$ is an odd integer, which implies that the order of $g = z^2 \bmod N$ is the same as the order of z, namely $\mathrm{ord}_N(g) = \mathrm{ord}_N(z) = \beta$. Also, since β is odd then $\mu \stackrel{\text{def}}{=} 2^{-1} \bmod \beta$ and $\tau \stackrel{\text{def}}{=} T^{-1} \bmod \beta$ exist and we have $z = g^\mu \bmod N$. Hence in this case we can write

$$x_0 = z \cdot g^{Tw} = g^{\mu + Tw} = g^{T(\tau\mu + w)} \mod N$$

Now notice that since w is random in $Z_{\varphi(N)}$ and β divides $\varphi(N)$, the random variables $\tau\mu+w \bmod \beta$ and $w \bmod \beta$ are identically distributed, so $x_0 = g^{T(w+\mu\tau)} \bmod N$ in the reduction is distributed identically to $x = g^{Tw} \bmod N$ in the protocol.

Case 2: $z \in J_N \setminus QR_N$**.** In this case we know that $z' \stackrel{\text{def}}{=} N - z$ is a quadratic residue modulo N. By the same reasoning as above, we know that $\beta \stackrel{\text{def}}{=} \text{ord}_N(z')$ is an odd integer, which implies that the order of $g = z^2 = (z')^2 \bmod N$ is also β, that $\mu \stackrel{\text{def}}{=} 2^{-1} \bmod \beta$ and $\tau \stackrel{\text{def}}{=} T^{-1} \bmod \beta$ exist, and that $z' = g^\mu \bmod N$. Hence in this case we can write

$$x_1 = -(z \cdot g^{Tw}) = z' \cdot g^{Tw} = g^{\mu+Tw} = g^{T(\tau\mu+w)} \mod N$$

and by the same arguments as above we have that $x_1 = g^{T(w+\mu\tau)} \bmod N$ in the reduction is distributed identically to $x = g^{Tw} \bmod N$ in the protocol. ■

Lemma 6 *The construction of $\mathcal{H}_{QR}$ from Section 6.2.1 is a verifiable-ϵ-universal projective hash family, where $\epsilon < \frac{1}{2} + O(2^{-n})$.*

Proof The projectiveness and the completeness of the instance-testing (conditions (a) and (c) in Definition 7) are easy to check: For any RSA modulus N and any $g \in R_N$, $w \in Z_N$, and $k \in Z_N$, setting $x = g^{Tw} \bmod N$ and $pk = (g^T)^k \bmod N$ (where $T = 2^{\lceil \log N \rceil}$) we get for Condition (a):

$$\mathsf{pHash}(N, g, pk, w, x) = pk^w = (g^{Tk})^w = (g^{Tw})^k = x^k = \mathsf{Hash}(N, g, k, x) \mod N$$

Also Condition (c) holds trivially.

The more interesting property is the verifiable-ϵ-universality (property (d')), which we prove next. (Note that properties (a), (c), and (d') together imply also property (b').) Let (N, g, x, x') be any four elements such that $\mathsf{IT}(N, g, x, x') = 1$. Namely, $N > 2^{2n}$, $g, x \in Z_N^*$, and $x' = N - x$. Note that since $x/x' = -1 \mod N$ and $\text{ord}_N(-1) = 2$, then at least one of x, x' has an even order modulo N. We next show that for any element z of even order, $\mathcal{H}_{QR}$ is ϵ-universal on (N, g, z), which implies that it must be ϵ-universal on at least one of x, x'.

Fix an odd modulus $N > 2$ and some $g, z \in Z_N^*$ such that $e \stackrel{\text{def}}{=} \text{ord}_N(z)$ is even. Also denote $\tau \stackrel{\text{def}}{=} \text{ord}_N(g^T)$ (where $T = 2^{\lceil \log N \rceil}$) and observe that τ must be odd and must divide $\varphi(N)$. We again make the simplifying assumption that the hashing key k is chosen from $Z_{\varphi(N)}$ instead of from Z_N, thus introducing an error of $O(1/\sqrt{N}) = O(2^{-n})$ into the analysis. Under this assumption we show that for any pk and any y, it holds that

$$\Pr_{k \in_R Z_{\varphi(N)}}[g^{Tk} = pk,\ y = z^k \mod N] \ \le\ \frac{1}{2} \cdot \Pr_{k \in_R Z_{\varphi(N)}}[g^{Tk} = pk \mod N]$$

We consider the following procedure for choosing $k \in_R Z_{\varphi(N)}$: First choose $k_0 \in_R \{0, \ldots, \tau - 1\}$, then $k_1 \in_R \{0, 1, \ldots, \frac{\varphi(N)}{\tau} - 1\}$, and then set $k \leftarrow k_0 + \tau \cdot k_1$. Note that $pk = (g^T)^k \mod N$ depends only on the choice of k_0. It is therefore sufficient to show that for any k_0 (that determines pk) and any y, it holds that $\Pr_{k_1}[z^{k_0+\tau \cdot k_1} = y \mod N] \le \frac{1}{2}$. In other words, let K_y be the set of all values that yield y, namely

$$K_y \stackrel{\text{def}}{=} \left\{ k_1 \in \left[0, \frac{\varphi(N)}{\tau} - 1\right] \ : \ z^{k_0+\tau \cdot k_1} = y \mod N \right\}.$$

We need to show that K_y contains no more than $\frac{\varphi(N)}{2\tau}$ values. We observe that e (the order of z) does not divide τ (the order of g^T) since e is even and τ is odd. Therefore $z^\tau \neq 1 \mod N$, which means that for any value of k_1, $z^{k_0+\tau\cdot k_1} \neq z^{k_0+\tau\cdot(k_1+1)} \mod N$. Hence the set K_y cannot contain two consecutive integers, and since $\frac{\varphi(N)}{\tau}$ is even then K_y cannot contain more than half the values in $[0, \frac{\varphi(N)}{\tau} - 1]$. This concludes the proof. ■

Acknowledgments. We would like to greatly thank Alon Rosen, whose comments, suggestions and involvement played an essential part in the creation of this work. We also thank Rosario Gennaro for many discussions on smooth projective hashing.

References

[AIR01] William Aiello, Yuval Ishai, Omer Reingold. Priced Oblivious Transfer: How to Sell Digital Goods. In *EUROCRYPT 2001*, pages 119-135, 2001.

[BS96] E.Bach and J. Shallit. Algorithmic Number Theory, Vol. 1: Efficient Algorithms. MIT Press, 1996.

[BM89] M. Bellare and S. Micali. Non-Interactive Oblivious Transfer and Applications. In *CRYPTO '89*, pages-547-557, 1989.

[Bl82] M. Blum. Coin flipping by telephone. In *Proc. IEEE Spring COMPCOM*, IEEE, 1982. Pages 133–137.

[CS03] J. Camenisch and V.Shoup. Practical Verifiable Encryption and Decryption of Discrete Logarithms. In *CRYPTO 2003*, Springer-Verlag (LNCS 2729), pages 126-144, 2003.

[CHK+05] R. Canetti, S.Halevi, J. Katz, Y. Lindell, P. MacKenzie. Universally Composable Password-Based Key Exchange. In *Eurocrypt 2005*, Springer-Verlag (LNCS 3494), pages 404-421, 2005.

[CS98] R. Cramer and V.Shoup. A Practical Public Key Cryptosystem Provably Secure Against Adaptive Chosen Ciphertext Attack. In *CRYPTO 1998*, pages 13-25, 1998.

[CS02] R. Cramer and V.Shoup. Universal Hash Proofs and a Paradigm for Adaptive Chosen Ciphertext Secure Public-Key Encryption. In *Eurocrypt 2002*, Springer-Verlag (LNCS 2332), pages 45-64, 2002.

[Cr87] C. Crépeau. Equivalence between two flavours of oblivious transfers. In *CRYPTO 1987*, pages 350-354, 1987.

[CCM98] C. Cachin, C. Crépeau, Julien Marcil. Oblivious Transfer with a Memory-Bounded Receiver. In *FOCS 1998*, pages 493-502, 1998.

[DHRS04] Y. Z. Ding, D. Harnik, A. Rosen, R. Shaltiel. Constant-Round Oblivious Transfer in the Bounded Storage Model. In *TCC 2004*, pages 446-472,2004.

[Fr98] John B. Fraleigh. A first course in abstract algebra , 7th edition, Addison-Wesley 1998.

[EGL85] S. Even and O. Goldreich and A. Lempel. A Randomized Protocol for Signing Contracts. In *Communications of the ACM 28:6*, pages 637–647, 1985.

[GL03] R. Gennaro and Y. Lindell. A Framework for Password-Based Authenticated Key Exchange. In *EUROCRYPT 2003*, pages 524-543, 2003.

[Go04] O. Goldreich. Foundations of Cryptography - Volume 2 (Basic Applications). Cambridge University Press, 2004.

[GMW87] O. Goldreich and S. Micali and A. Wigderson. How to Play any Mental Game - A completeness Theorem for Protocols with Honest Majority. In *STOC 1987*, pages 218-229, 1987.

[Hai04] Iftach Haitner. Implementing Oblivious Transfer Using Collection of Dense Trapdoor Permutations. In *TCC 2004*, pages 394-409, 2004.

[Hal99] Shai Halevi. Efficient Commitment Schemes with Bounded Sender and Unbounded Receiver. In *J. Cryptology* 12(2): 77-89 (1999).

[IR89] R. Impagliazzo and S. Rudich. Limits on the Provable Consequences of One-Way Permutations. In *STOC 89*, pages 44–61, 1989.

[KOY01] J. Katz, R. Ostrovsky, M. Yung. Efficient Password-Authenticated Key Exchange Using Human-Memorable Passwords. In *EUROCRYPT 2001*, pages 475-494, 2001.

[Ki88] J. Kilian. Founding Cryptography on Oblivious Transfer. 20th ACM Symposium on the Theory of Computing, pages 20–31, 1988.

[NP01] M. Naor and B. Pinkas. Efficient oblivious transfer protocols. In *SODA 2001*, pages 448-457,2001

[NZ96] N. Nissan and D. Zuckerman. Randomness in linear space. *Journal of Computer and System Science 52(1): 43–52*, February 1996.

[Pa99] P. Paillier. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In *EUROCRYPT 1999*, pages 223-238, 1999.

[Rab81] M. O. Rabin. How to Exchange Secrets by Oblivious Transfer. TR-81, Harvard, 1981.

[Y86] A. C. Yao. How to Generate and Exchange Secrets. In *FOCS 1986*, pages 162-167, 1986.

A Factoring-Based CCA-Secure Encryption without Safe Primes

In this section we show that the factoring-based CCA secure encryption schemes from [CS02, GL03, CS03] — all of which were only proven secure when used with RSA-composites that are product of safe primes — remain secure even when used with "non-safe" RSA composites. These schemes are all very similar and all are proved in more-or-less the same way. Below we first give a very high level explanation of how to modify the security proofs from [CS02, GL03] to get security even when used with "non-safe" RSA composites. We then give a detailed description of how to modify the proof from [CS03] to get security even when used with "non-safe" RSA composites.

A.1 Modifying the Proofs: High Level Description

We begin with a very high-level description of the Cramer-Shoup construction of a CCA-secure encryption scheme from a smooth projective hash family. (This description is over-simplified and contains only the details that are important to convey our ideas.)

The encryption scheme. Very loosely speaking, the key generation algorithm first runs the parameter-generator algorithm PG to obtain parameters $\Lambda \leftarrow \mathsf{PG}(1^n)$. Then it runs the hash-key generator algorithm HG to obtain a primary hashing key and a projective key $(hk, pk) \leftarrow \mathsf{HG}(\Lambda)$. It outputs (Λ, pk) as the public key, and hk as the secret key. The parameters Λ define a hashing domain X and a "special subset" L. The encryption algorithm chooses an element x in the special subset L (together with a corresponding witness), and computes a tag $\hat{\pi}$ which is (essentially) the hash of x, computed using the projective key pk included in the public key. The ciphertext is a triple $(x, \hat{\pi}, e)$, where e is the element that actually carries information about the encrypted message (but we mostly ignore the element e in this high-level description). The decryption algorithm uses the corresponding primary hashing key hk included in the secret key to re-compute the hash of x. It rejects the ciphertext if the computed hash differs from $\hat{\pi}$, and otherwise it extracts the message from e. The proof of CCA-security then uses the following arguments:

1. The answers to decryption queries, where the x component belongs to the special subset L, do not give any more information about the primary hashing key beyond what's implied by the public key itself. (This is an information-theoretic argument that follows from the projectiveness of the hashing scheme.)
2. Decryption queries where x is not in the special subset L (i.e., where $x \in X \setminus L$), have incorrect tags w.h.p. (This is an information-theoretic argument that follows from the smoothness of the hashing scheme.)
3. The adversary cannot distinguish whether the x component of the challenge ciphertext is in the special subset L or is in $X \setminus L$ (because of the hard subset membership property).
4. If the x component of the target ciphertext is in $X \setminus L$, then the target ciphertext contains no information about the plaintext, assuming no additional information about the primary hashing key hk is given beyond its corresponding projective key pk. (This is again an information-theoretic argument, similar to the argument about the smoothness of the hashing scheme.)

Roughly speaking, the last two arguments imply that barring some additional information from the decryption queries, the target ciphertext cannot be distinguished from an entity that carries no information about the plaintext. The first two arguments, on the other hand, imply that the attacker cannot use the decryption queries to get any information beyond what's already implied by the public key. Combining these arguments, CCA security follows.

The difficulty in removing the "safe prime requirement." Trying to use the constructions from Section 6 to get CCA security as above we run into some issues. Consider for example the construction of projective hashing based on the N'th Residuosity Assumption. In this case, the parameters are $\Lambda = (N, g)$ where N is an RSA composite and g is an N'th residue modulo N^2, the "special subset" is $L = \langle g \rangle$, and the witness for membership in L is the exponent r such that

$g^r = x \bmod N^2$. When N is not a product of safe primes, the set of N'th residues in $Z^*_{N^2}$ is not a cyclic group (even when restricted to elements with Jacobi symbol $+1$), and in particular the "special subset" $L = \langle g \rangle$ does not contain all the N'th residues in $Z^*_{N^2}$. Since the hash function is only smooth on non-N'th-residues, we can no longer claim that it is smooth on any $x \notin L$.

In the Oblivious-Transfer application we "solved" this problem by defining $X = \langle g \rangle \cdot \langle 1+N \rangle$, which ensures that every element in $X \setminus L$ is a non-N'th-residue. However, in a CCA attack the attacker may choose to supply ciphertext elements that are not in this set X, and there does not seem to be an easy way of deciding whether or not $x \in X$. In essence, our problem is that we cannot rule out the possibility that the attacker may find ciphertext elements which are neither in the smooth domain nor in the projective domain.

Our solution. We overcome this problem by considering four sets rather than two: Namely, we have the "big domain" $X = Z^*_{N^2}$, the "small domain" $X^* = \langle g \rangle \cdot \langle 1+N \rangle$, the "projective subset" $L = R_N$ (of all the N'th residues modulo N^2) and the "special subset" $L^* = \langle g \rangle = X^* \cap L$. The encryption scheme itself would produce only elements that belong to the "special subset" L^*, but the attacker can submit decryption queries with arbitrary elements from the "big domain" X.

Then we use the facts that (a) the hashing scheme is projective on the "projective subset" L,[11] (b) the hashing scheme is smooth (or at least universal) on $X \setminus L$, and (c) the uniform distributions on L^* and X^* are indistinguishable (this is Corollary 3). The four arguments from above are refined as follows:

1′. The answers to decryption queries, where the x component belongs to the projective subset L, do not give any more information about the primary hashing key beyond what's implied by the public key itself.[11] (This is an information-theoretic argument that follows from the protectiveness of the hashing scheme.)

2′. Decryption queries where x is in $X \setminus L$ have incorrect tags w.h.p. (This is an information-theoretic argument that follows from the smoothness of the hashing scheme.)

3′. The adversary cannot distinguish whether the x component of the challenge ciphertext is in the special subset L^* or is in $X^* \setminus L^*$ (because of the hard subset membership property).

4′. If the x component of the target ciphertext is in $X^* \setminus L^*$ then the target ciphertext contains no information about the plaintext, assuming no additional information about the primary hashing key hk is given beyond its corresponding projective key pk. (This is again an information-theoretic argument, similar to the argument about the smoothness of the hashing scheme.)

As before, arguments 3′ and 4′ imply that barring some additional information from the decryption queries, the target ciphertext cannot be distinguished from an entity that carries no information about the plaintext. Arguments 1′ and 2′, on the other hand, imply that the attacker cannot use the decryption queries to get information that will help him distinguish the target ciphertext from an entity that carries no information about the plaintext.

Remark. Argument 1′ from above is slightly incorrect, in that the hashing schemes from Section 6 may fail to be projective on the "projective subset" L. Specifically, if the order of the element g

[11] This statement is not precise; see remark below.

(which is a random N'th residue) is not maximal (among the N'th residues), then the projective key g^{hk} does not contain enough information on the hashing key hk to fully determine the value of x^{hk} for every N'th residue x.

An easy solution here is to consider a mental experiment in which the hashing scheme is augmented by including a maximal-order N'th residue h as a parameter and h^{hk} as part of the projective key (so both h, h^{hk} are included in the public key of the encryption scheme). These elements are never used by the encryption or decryption algorithms; their sole purpose is to leak to the adversary more information on the secret key, so as to make the Argument $1'$ from above correct.

A.2 Reconstructing the Camenisch-Shoup Proof

We next reconstruct the proof from (the full version of) [CS03], augmenting it to show that the N'th-Residuosity-based encryption scheme of Gennaro-Lindell/Camenisch-Shoup [GL03, CS03] is CCA secure even when the modulus is not a product of safe primes.[12] This modification is based on the same approach as in Section A.1, but is explained in greater detail. The encryption scheme consists of the following three algorithms.

Key generation. On security parameter n, choose a seed s for a collision-resistant hash function H_s, choose $p, q \in_R \mathsf{Primes}(n)$, and set $N \leftarrow pq$. Next choose at random $g' \in_R Z^*_{N^2}$ and three integers $x_1, x_2, x_3 \in_R [N^2/4]$, and set $g \leftarrow (g')^{2N}$, $y_1 \leftarrow g^{x_1}$, $y_2 \leftarrow g^{x_2}$, and $y_3 \leftarrow g^{x_3}$ (all modulo N^2). Output (s, N, g, y_1, y_2, y_3) as the public key, and (x_1, x_2, x_3) as the secret key.

Encryption. To encrypt $m \in [N]$ (with label $L \in \{0,1\}^*$) choose at random $r \in_R [N/4]$ and set

$$e \leftarrow y_1^r(1+N)^m, \quad u \leftarrow g^r, \quad \alpha \leftarrow H_s(u, e, L), \quad v \leftarrow \mathsf{abs}\left((y_2 y_3^\alpha)^r\right),$$

where all the calculation are modulo N^2, and $\mathsf{abs}(x) \stackrel{\text{def}}{=} \min\{x, N^2 - x\}$. The ciphertext is (e, u, v).

Decryption. To decrypt the ciphertext (e, u, v) with label L, set $\alpha \leftarrow H_s(u, e, L)$, and check that $v = \mathsf{abs}(v)$ and $u^{2(x_2+\alpha x_3)} = v^2 \pmod{N^2}$. If so, let $z \leftarrow (e/u^{x_1})^{N+1} \pmod{N^2}$, and if $z-1$ is divisible by N (over the integers) then compute $m \leftarrow \frac{z-1}{N}$ and output m.

Remark. Some aspects of this construction are not really relevant for our discussion: These include the squaring of g' during key-setup, the use of the "absolute values", the squaring of u and v and exponentiation to the $N+1$ power during decryption, and also the inclusion of the attached labels. The reason that we keep these aspects here is because they are present in the scheme and proof from [CS03], and we want to stress that we are proving the exact same scheme (using an almost identical proof), except that we omit the requirement of using safe primes. However, it is clear that the same proof as below can be applied also to the simplified scheme without these components.

A.2.1 Proof of security

Lemma 7 *Under the N'th Residuosity Assumption, the encryption scheme from Section A.2 is CCA secure.*

[12] We chose to reconstruct the proof from [CS03] since the presentation of that proof is more "from first principles" than that of the proof from [GL03].

Proof The proof mimics closely the proof of Theorem 1 from the full version of [CS03] by describing almost the exact same nine games that were considered in the proof of Theorem 1 there. In fact, the only steps in which our proof differs from the one in [CS03] are Games 4-5 (see below).

Game 0. This is the standard CCA-game. Namely, the key-generation algorithm is run on security parameter n, resulting in public key (s, N, g, y_1, y_2, y_3) and secret key (x_1, x_2, x_3) as above. The attacker gets the public key as input, and it may issue decryption queries (e_i, u_i, v_i, L_i), $i = 1, 2, \ldots$ For each query, the attacker gets the result of running the decryption algorithm on these ciphertexts and labels. (This is called "Probing phase I").

Next the attacker outputs two messages $m_0, m_1 \in [N]$ and label L^*. Then a random bit $\sigma \in_R \{0, 1\}$ is chosen and the message m_σ is encrypted (with label L^*). This is done by choosing $r^* \in_R [N^2/4]$ and setting

$$u^* \leftarrow g^{r^*}, \quad e^* \leftarrow y_1^{r^*}(1+N)^{m_\sigma}, \quad \alpha^* \leftarrow H_s(e^*, u^*, L^*), \text{ and } v^* \leftarrow \mathsf{abs}\left((y_2 y_3^{\alpha^*})^{r^*}\right).$$

Below we call m_0, m_1 the "target messages" and (e^*, u^*, v^*) the "target ciphertext". The target ciphertext is returned to the attacker, and then the attacker can keep making decryption queries as before, under the condition that $(e_i, u_i, v_i, L_i) \neq (e^*, u^*, v^*, L^*)$. (This is called "Probing phase II".) Finally the attacker outputs a bit σ', and it is considered successful if and only if $\sigma' = \sigma$.

The goal of the analysis is to prove that under the N'th Residuosity Assumption, the attacker cannot succeed with probability noticeably better than 1/2. The analysis proceeds by making successive small changes to the way some variables are computed in this game, each time proving that the change can have at most a negligible effect on the success probability of the attacker, until arriving at a game where the attacker's view is independent of the bit σ (and therefore its success probability is exactly 1/2).

Game 1. The only difference between this game and the previous one is that the decryption oracle rejects any ciphertext query (e_i, u_i, v_i, L_i) during "Probing phase II" such that $(e_i, u_i, L_i) \neq (e^*, u^*, L^*)$ but $H_s(e_i, u_i, L_i) = H_s(e^*, u^*, L^*)$. Clearly this only happens if the attacker finds a collision in the hash function H_s, so the success probability in this game is at most negligibly different than in Game 0.

Game 2. Next we also reject ciphertext queries (e_i, u_i, v_i, L_i) during "Probing phase II" such that $v_i \neq v^*$ but $v_i^2 = (v^*)^2$. Observe that since $v_i, v^* < N^2/2$ then the condition above implies finding a nontrivial square root of unity, and hence factoring N. It follows that this modification too can only change the success probability by a negligible amount.

Game 3. We now change the way the target ciphertext is computed. Specifically, we now compute

$$e^* \leftarrow (u^*)^{x_1}(1+N)^{m_\sigma}, \text{ and } v^* \leftarrow \mathsf{abs}\left((u^*)^{x_2+\alpha^* x_3}\right)$$

(where $\alpha^* = H_s(u^*, e^*, L^*)$). As these values coincide with the values of e^*, v^* that were computed before, this modification has no effect on the success probability.

Games 4-5. *This is the only difference between our proof the the one from [CS03].* In the proof from [CS03], Game 4 modifies the choice of u^*, choosing it as a random square in the set of N'th residues modulo N^2 (instead of setting $u^* \leftarrow g^{r^*}$ for a random exponent $r^* \in_R [N^2/4]$), and Game 5 modified this choice again, choosing u^* as a random square in $Z^*_{N^2}$.

In our proof, we skip Game 4 altogether and in Game 5 we choose u^* as a random element in $\langle g \rangle \cdot \langle 1+N \rangle$. Here we appeal to Corollary 3, which tells us that for an N'th residue g, a random element in $\langle g \rangle$ is indistinguishable from a random element in $\langle g \rangle \cdot \langle 1+N \rangle$. Hence the difference in success probability between Game 3 and Game 5 must be negligible. (Note that in Corollary 3 g was chosen as a random N'th residue, whereas here g is chosen as a random $2N$'th residue. However, it is clear that if $\langle \mu^N \rangle$ is indistinguishable from $\langle \mu^N \rangle \cdot \langle 1+N \rangle$ then also $\langle \mu^{2N} \rangle$ is indistinguishable from $\langle \mu^{2N} \rangle \cdot \langle 1+N \rangle$.)

Game 6. As done in the proof from [CS03], we now choose u^* not as a completely random element in $\langle g \rangle \cdot \langle 1+N \rangle$, but rather a random element in that group subject to the restriction that its order is divisible by N. Since a random element in that group satisfies this restriction with all but exponentially small probability, then this has almost no effect on the success probability.

Game 7. We now modify the key-generation algorithm, choosing $x_1, x_2, x_3 \in_R [N \cdot \varphi(N)/4]$ (instead of choosing them from $[N^2/4]$). This only has an exponentially small effect on the success probability.

Game 8. The last game modifies again the decryption oracle, this time rejecting any ciphertext query (e_i, u_i, v_i, L_i) for which u_i is not an N'th residue modulo N^2.

Denote $\lambda(N) \stackrel{\text{def}}{=} \varphi(N)/4$ and let $x_1' = x_1 \bmod \lambda(N)$ and $x_1'' = x_1 \bmod N$. It is fairly easy to see that in this game, the public key and the answers of the decryption oracle are independent of x_1'' and depend on x_1 only through x_1' (recall that g is a $2N$'th residue, so its order divides $\lambda(N)$). Moreover we know that $x_1'' \in_R [N]$, and we know that the order of u^* is divisible by N (and u^* is a square), which means that $u^* = w(1+N)^t$ with w a $2N$'th residue and t co-prime with N. Since the element e^* is computed as

$$e^* \leftarrow (u^*)^{x_1}(1+N)^{m_\sigma} = w^{x_1'}(1+N)^{t \cdot x_1'' + m_\sigma},$$

then the distribution of e^* is independent of m_σ (and thus also of σ). It follows that the view of the attacker is independent of σ, and therefore its success probability in this game is exactly $1/2$.

It remains to bound the difference between the success probability of the attacker in games 7 and 8. Namely, we need to bound the probability that there exists some decryption query (e_i, u_i, v_i, L_i) in Game 8 such that $v_i = \mathsf{abs}(v_i)$, $u_i^{2(x_2+\alpha_i x_3)} = v_i^2$, the two conditions from Games 1 and 2 do not hold, and yet u_i is not an N'th residue modulo N^2.

Consider a particular decryption query (e_i, u_i, v_i, L_i) for which u_i is not an N'th residue modulo N^2, and denote by o_i the order of u_i in $Z^*_{N^2}$. We first observe that *o_i is not co-prime with N*. Indeed, if o_i were co-prime with N then there would exist integers a, b such that $aN + bo_i = 1$, and therefore

$$(u_i^a)^N = (u_i^a)^N (u_i^{o_i})^b = u_i^{aN+bo_i} = u_i \mod N^2,$$

contradicting our working assumption that u_i is not an N'th residue modulo N^2. Since $N = pq$ with p, q primes, it follows that the order of u_i is divisible by either p or q (or both).

The rest of the argument follows the exact same line as in the proof of [CS03] (but our presentation is slightly different). We observe that the view of the attacker is completely determined by the following values:

- N, g and $x_i \bmod \lambda(N)$ (which completely determine the answers of all the decryption queries),
- u^*, σ and $x_1 \bmod N$ (which together with the values above determine the value of the element e^* of the target ciphertext), and

- $x_2 + \alpha^* x_3 \bmod N$ (which together with the values above determines the value of the element v^* of the target ciphertext.

We therefore consider the alternative view of Game 8 where the values N, g, $x_i \bmod \lambda(N)$, u^*, σ, and $x_1 \bmod N$ are chosen at the outset, and the values of x_2, x_3 are chosen as follows:

- If the i'th decryption query was made during "Probe phase I" then we choose $x_2, x_3 \bmod N$ after the attacker makes this query. Since both x_2, x_3 are uniform in $[N]$ and since u_i, v_i are fixed and the order of u_i is divisible by p or q (and therefore so is the order of u_i^2), then the probability of getting $u_i^{2(x_2+\alpha_i x_3)} = v_i^2$ is at most $1/\min(p, q)$.
- If the i'th decryption query was made during "Probe phase II" then we choose the value of $x_2 + \alpha^* x_3 \bmod N$ after the attacker determines the target messages m_0, m_1, and we choose $x_2 + \alpha_i x_3 \bmod N$ after the attacker makes the i'th decryption query. Since $\alpha_i \neq \alpha^*$ (and they are both smaller than N) then the value of $x_2 + \alpha_i x_3 \bmod N$ is still uniform in $[N]$ even after $x_2 + \alpha^* x_3 \bmod N$ is fixed. Again, this implies that the probability of getting $u_i^{2(x_2+\alpha_i x_3)} = v_i^2$ is at most $1/\min(p, q)$.

We therefore determine that the probability that any decryption query i induces a difference between Game 7 and Game 8 is at most $1/\min(p, q)$, and therefore the difference in the success probability between these two games is at most $\kappa/\min(p, q)$ where κ is the number of decryption queries. This completes the security proof. ∎