Preprint, November 16, 2015.

A

A New Framework for Privacy-Preserving Aggregation of Time-Series Data

FABRICE BENHAMOUDA, ENS, CNRS, INRIA, and PSL, Paris, France
MARC JOYE, Technicolor, Los Altos, CA, USA
BENOÎT LIBERT, ENS Lyon, Lyon, France

Aggregator-oblivious encryption is a useful notion put forward by Shi *et al.* in 2011 that allows an untrusted aggregator to periodically compute an aggregate value over encrypted data contributed by a set of users. Such encryption schemes find numerous applications, in particular in the context of privacy-preserving smart metering.

This paper presents a general framework for constructing privacy-preserving aggregator-oblivious encryption schemes using a variant of Cramer-Shoup's paradigm of smooth projective hashing. This abstraction leads to new schemes based on a variety of complexity assumptions. It also improves upon existing constructions, providing schemes with shorter ciphertexts and better encryption times.

General Terms: Data aggregation, Privacy, Cryptography

Additional Key Words and Phrases: Private aggregation, aggregator-oblivious encryption, smart metering, smooth projective hashing, security reductions

1. INTRODUCTION

Since the introduction of electricity into the home, its usage has been recorded by an electricity meter attached to the exterior of the homes. This situation is however gradually changing with the progressive deployment of smart meters [McDaniel and McLaughlin 2009]. In addition to the basic service offered by its predecessor, a smart meter comes with extra useful features aiming at reducing energy costs. For example, a smart meter can turn down momentarily high-energy electrical appliances during peak hours. For the utility company, one of its most appealing features is its ability to report in almost real-time the power consumption of their consumers. This fine-grained information is very helpful as it allows the electricity provider to better adapt the load or forecast the supply. Moreover, it allows the electricity provider to quickly react when anomalies are detected on the grid. The resulting savings are also beneficial to the consumers as they give rise to better pricing. But there is a downside. Frequent usage reports leak information about the consumer habits and behaviors —for example, when a certain consumer turns her TV on and what programs she is likely to watch. These seemingly unimportant issues should not be underestimated as they may have unintended consequences from the inference of some private information (attributes or data).

In most cases, there is no need for the utility company (except for preparing the monthly bill) to get the fined-grained energy usage of *each* customer. For example, in the previous scenario, an aggregate suffices. The goal of this paper is to mitigate the privacy issues that arise from smart metering by computing aggregates rather than individual energy consumption. More generally, we are seeking efficient privacy-preserving methods for the aggregation of time-series data. The entity computing the aggregates is not necessarily trusted. We are interested in solutions that affect the existing infrastructure as little as possible. In particular, in the above scenario, we do not require smart meters to interact with each other nor the existence of a return channel. We note that the billing issue is separate. In practice, smart meters report separately their monthly energy consumption to the energy provider.

Related work. The above setting is the one considered in [Shi et al. 2011] and [Joye and Libert 2013]. Each smart meter encrypts its actual energy consumption and sends the result at regular intervals to an aggregator (that can be an entity different from the

energy provider). In the case of electricity metering, a typical time period is 15 minutes. Upon receiving the encrypted values from a predetermined set of users, the aggregator combines the received values and therefrom deduces the total energy consumption over the population of these users for the current time period. This operation involves a secret key known only to the aggregator. Further, computing the sum over the predetermined set of users is the only operation the aggregator can perform —it cannot learn anything beyond what is revealed by the aggregate value. Following [Shi et al. 2011], such a scheme is termed an *aggregator-oblivious encryption scheme*.

Like [Shi et al. 2011] and [Joye and Libert 2013], all our schemes can serve as a building block for the fault tolerant solution of [Chan et al. 2012] while enjoying the benefits of our construction. In fact, all the extensions of [Shi et al. 2011] are also possible with our system. In particular, although the focus of this paper is put on the encryption, the proposed schemes are compatible with the differential-privacy framework [Dwork 2008; Dwork et al. 2006]. In this case, the smart meters simply need to add some appropriately-generated noise to their data prior to encryption

Two other protocols in settings similar to the one of aggregator-oblivious encryption are the low-overhead protocol of [Kursawe et al. 2011] and the protocol of [Ács and Castelluccia 2011]. These protocols however have the drawback of requiring each smart meter to store as many keys as there are users, which can be impractical for a large set of smart meters.

We also note that recent results on multi-input functional encryption [Goldwasser et al. 2014] imply non-interactive constructions of aggregator-oblivious encryption. However, due to their inevitable use of indistinguishability obfuscation candidates [Garg et al. 2013], they are really far from being practical and should only be seen as feasibility results.

Many other settings than the one we consider have been studied in the literature. For example, the protocol of Dwork *et al.* in [Dwork et al. 2006] allows the aggregation of more complex functions, but requires communication between the smart meters. The setting of [Rastogi and Nath 2010] and of [Garcia and Jacobs 2010] require bi-directional channels between the smart meters and the aggregator, while we only require a uni-directional channel from each smart meter to the aggregator. The protocols of [Leontiadis et al. 2014] as well as [Jawurek and Kerschbaum 2012] suppose the existence of an additional semi-trusted party, who cannot collude with the aggregator. In addition, in the second paper, the smart meter should be able to receive data from this third party.

For more information on aggregation schemes, we refer the reader to the detailed survey of Jawurek, Kerschbaum, and Danezis [Jawurek et al. 2012].

Our contributions. While applicable to our framework, the solutions offered in [Shi et al. 2011] and [Joye and Libert 2013] are not fully satisfying, but for different reasons. Table I gives a rough idea of the expected gains for our basic aggregator-oblivious encryption scheme (a more detailed analysis with concrete implementation numbers is provided in Section 3).

Joye-Libert's scheme supports large plaintext spaces and a large number of users. However as it is built over Paillier's encryption scheme, the involved parameters are somewhat large. Back to our example of smart meters, this in turn implies that these are likely equipped with crypto-processors for modular arithmetic over large integers and possess sufficient memory for storing intermediate computations. Larger ciphertexts also mean more bandwidth for their transmission. Shi *et al.*'s scheme provides a cheaper solution as it is ElGamal-based and relies on the Decisional Diffie-Hellman assumption (DDH). In particular, it can be implemented using elliptic-curve groups with much shorter parameters. It requires the computed aggregated sum to lie in a relatively small predetermined set. In the case of smart meters, this does not really constitute a limitation for most practical settings, as the sum should be less than 30 bits long.

A reductionist security proof of a cryptographic scheme consists in an efficient algorithm, called a reduction, that uses an attacker against the scheme as a subroutine to solve a

Table I. Reduction loss and typical parameter size for existing schemes and our basic scheme (for $T = n = 2^{20}$, based on ECRYPT 2 recommendations, $s^{-1}(\alpha)$ is the minimal number of bits of a modulus which cannot be factored in time 2^{α}, see Section 3.3 for details)

Security level	Scheme	Reduction loss	Typical size (bits)		
			Group elements	Private keys	Ciphertexts
80 bits	[Shi et al. 2011]	$\approx 2^{80}$	320	320	320
80 bits	[Joye and Libert 2013]	$\lesssim 2^{20}$	≤ 3862	≤ 3862	≤ 3862
80 bits	**This work**	$\lesssim \mathbf{2^{20}}$	$\leq$ **200**	$\leq$ **400**	$\leq$ **200**
128 bits	[Shi et al. 2011]	$\approx 2^{80}$	416	416	416
128 bits	[Joye and Libert 2013]	$\lesssim 2^{20}$	≤ 8900	≤ 8900	≤ 8900
128 bits	**This work**	$\lesssim \mathbf{2^{20}}$	$\leq$ **296**	$\leq$ **592**	$\leq$ **296**
λ bits	[Shi et al. 2011]	$\approx Tn^3$	$2(\lambda + \log_2(Tn^3))$	$2(\lambda + \log_2(Tn^3))$	$2(\lambda + \log_2(Tn^3))$
λ bits	[Joye and Libert 2013]	$\lesssim T$	$\leq 2s^{-1}(\lambda + \log_2(T))$	$\leq 2s^{-1}(\lambda + \log_2(T))$	$\leq 2s^{-1}(\lambda + \log_2(T))$
$\boldsymbol{\lambda}$ **bits**	**This work**	$\lesssim \boldsymbol{T}$	$\leq \mathbf{2(\lambda + \log_2(T))}$	$\leq \mathbf{4(\lambda + \log_2(T))}$	$\leq \mathbf{2(\lambda + \log_2(T))}$

problem supposed hard. If ε and ε' respectively denote the success probability of the attacker and of the reduction algorithm, the security proof is said *tight* when $\varepsilon \approx \varepsilon'$ and *loose* otherwise. The tightness gap is measured by the ratio ε/ε' and captures the security loss. This ratio is an important parameter as it defines the *exact security* [Bellare and Rogaway 1996; Bellare 1998] of a scheme. It quantifies the amount by which the security parameters defining the scheme need to be increased to accommodate the tightness gap.

But as already pointed out in [Joye and Libert 2013], one drawback of [Shi et al. 2011] is that the security reduction from the underlying complexity assumption is very loose. If the scheme is set up for n users and if the number of time periods is at most T, there is a multiplicative gap of $O(Tn^3)$ between the adversary's advantage and the reduction's probability to solve the DDH problem. Moreover, using a meta-reduction, we show in Appendix C that a degradation factor of at least $\Omega(n^2)$ is unavoidable in the scheme of [Shi et al. 2011].

An important contribution of this paper is a new DDH-based aggregator oblivious encryption scheme (cf. Section 3) with a much tighter security reduction. While the security loss is $O(Tn^3)$ in Shi *et al.*'s scheme, our basic scheme reduces this gap to roughly $O(T)$ in the worst-case scenario (this worst-case scenario is illustrated by the figures given in Table I).

In Section 4, we generalize our basic DDH-based construction. We propose a generic framework for the privacy-preserving aggregation of time-series data featuring a tighter reduction. This framework is based on smooth projective hash functions [Cramer and Shoup 2002] (SPHFs) with an extra additively homomorphic property over the key space. Note that all SPHF realizations given in [Cramer and Shoup 2002] are natively additively key-homomorphic. As shown in Section 5, our framework encompasses our basic scheme as well as a variation of Joye-Libert's scheme. Several other aggregator-oblivious encryption schemes based on a variety of complexity assumptions are presented in Section 5. This clearly demonstrates the generic aspect of our framework.

2. AGGREGATOR-OBLIVIOUS ENCRYPTION

We review the definition of aggregator-oblivious (secret-key) encryption and then proceed with the corresponding security notion. We refer the reader to [Shi et al. 2011] for further introductory background.

Definition 2.1. An *aggregator-oblivious encryption scheme* is a tuple of three algorithms, (Setup, Enc, AggrDec), defined as:

$\mathsf{Setup}(1^\lambda)$. On input a security parameter λ, a trusted dealer generates the system parameters param, the aggregator's private key sk_0, and the private key sk_i for each user i $(1 \leq i \leq n)$.
$\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, \tau, x_{i,\tau})$. At time period τ, user i encrypts a value $x_{i,\tau}$ using her private encryption key sk_i to get $c_{i,\tau} = \mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, \tau, x_{i,\tau})$.
$\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, \tau, c_{1,\tau}, \ldots, c_{n,\tau})$. At time period τ, the aggregator using sk_0 obtains

$$X_\tau = \sum_{i=1}^{n} x_{i,\tau} \bmod M\,,$$

as the evaluation of $X_\tau = \mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, \tau, c_{1,\tau}, \ldots, c_{n,\tau})$. M is some fixed integer contained in the system parameters param.

This definition slightly generalizes the definition introduced in [Shi et al. 2011]. We make explicit the fact the sum is computed modulo some integer M. To compute sums over the integers, it is sufficient to choose M greater than the maximal possible sum, as done in the constructions of [Shi et al. 2011]. Furthermore, we note that some constructions assume $\mathsf{AggrDec}$ only works on a small subset of $\{0, \ldots, M-1\}$.

2.1. Aggregator obliviousness

Basically, the security notion of *aggregator obliviousness* (AO) requires that the aggregator cannot learn, for each time period, anything more than the aggregate value X_τ from the encrypted values of n (honest) users. If there are corrupted users (i.e., users sharing their private information with the aggregator), the notion only requires that the aggregator gets no extra information about the values of the honest users beyond their aggregate value. Further, it is assumed that each user encrypts only one value per time period.

More formally, AO is defined by the following game between a challenger and an attacker.

Setup. The challenger runs the Setup algorithm and gives param to the attacker.
Queries. In a first phase, the attacker can submit queries that are answered by the challenger. The attacker can make two types of queries:
(1) Encryption queries: The attacker submits tuples $(i, \tau, x_{i,\tau})$ for a pair (i, τ) and gets back the encryption of $x_{i,\tau}$ under key sk_i for time period τ;
(2) Compromise queries: The attacker submits i and receives the private key sk_i of user i; if $i = 0$, the attacker receives the private key of the aggregator.
Challenge. In a second phase, the attacker chooses a time period $\tau^\star$. Let $\mathcal{U}^\star \subseteq \{1, \ldots, n\}$ be the whole set of users for which, at the end of the game, no encryption queries have been made on time period $\tau^\star$ and no compromise queries have been made. The attacker chooses a subset $\mathcal{S}^\star \subseteq \mathcal{U}^\star$ and two different series of triples

$$\langle (i, \tau^\star, x^{(0)}_{i,\tau^\star}) \rangle_{i \in \mathcal{S}^\star} \quad \text{and} \quad \langle (i, \tau^\star, x^{(1)}_{i,\tau^\star}) \rangle_{i \in \mathcal{S}^\star}\,,$$

that are given to the challenger. Further, if the aggregator capability sk_0 is compromised at the end of the game and $\mathcal{S}^\star = \mathcal{U}^\star$, it is required that

$$\sum_{i \in \mathcal{S}^\star} x^{(0)}_{i,\tau^\star} \bmod M = \sum_{i \in \mathcal{S}^\star} x^{(1)}_{i,\tau^\star} \bmod M\ . \tag{1}$$

Guess. The challenger chooses uniformly at random a bit $b \in \{0, 1\}$ and returns the encryption of $\langle x^{(b)}_{i,\tau^\star} \rangle_{i \in \mathcal{S}^\star}$ to the attacker.
More queries. The attacker can make more encryption and compromise queries. Note that since $\mathcal{S}^\star \subseteq \mathcal{U}^\star$, the attacker cannot submit an encryption query $(i, \tau^\star, \cdot)$ with $i \in \mathcal{S}^\star$ or a compromise query i with $i \in \mathcal{S}^\star$.

Outcome. At the end of the game, the attacker outputs a bit b' and wins the game if and only if $b' = b$.

Definition 2.2. An encryption scheme is said to meet the AO *security notion* if no probabilistic polynomial-time attacker can guess correctly, in the above game, the bit b with a probability non-negligibly better (in the security parameter) than $1/2$. Formally, an encryption scheme is AO-*secure* if the advantage of any probabilistic polynomial-time attacker $\mathcal{A}$ defined as

$$\mathbf{Adv}^{\mathsf{AO}}(\mathcal{A}) := 2\left|\Pr[b' = b] - 1/2\right|$$

is negligible; the probability is taken over the random coins of the game according to the distribution induced by Setup and over the random coins of the attacker.

2.2. Existing schemes

So far, there are two known constructions of AO encryption schemes. They both meet the AO security notion, in the random oracle model. The first one is due to Shi, Chan, Rieffel, Chow and Song [2011] and works in DDH groups. The second construction, due to Joye and Libert [2013], relies on the composite residuosity assumption [Paillier 1999]. These two schemes are reviewed in Appendix A.

3. A NEW DDH-BASED SCHEME

As aforementioned, the security proof offered in [Shi et al. 2011] incurs an $O(Tn^3)$ degradation factor. The scheme in [Joye and Libert 2013] avoids this degradation factor; namely, the multiplicative gap between the adversary's maximal advantage and the probability to break the underlying complexity assumption is only proportional to the number q_{enc} of encryption queries made by the adversary for distinct time periods other than $\tau^\star$ (so that, $q_{enc} < T$). In this section, we introduce an aggregator-oblivious encryption scheme enjoying a security reduction as tight as in [Joye and Libert 2013] but based on the DDH assumption. The main advantage is that the resulting ciphertexts are much shorter.

3.1. Basic scheme

We base the security of our basic scheme on the standard DDH assumption.

Definition 3.1. Let $\mathbb{G}$ be a group of prime order p. The *Decision Diffie-Hellman* (DDH) problem in $\mathbb{G}$ is to distinguish among the following two distributions:

$$D_0 = \left\{(g, g^a, g^b, g^{ab}) \mid g \xleftarrow{R} \mathbb{G},\ a, b \xleftarrow{R} \mathbb{Z}_p\right\}$$

and

$$D_1 = \left\{(g, g^a, g^b, g^c) \mid g \xleftarrow{R} \mathbb{G},\ a, b, c \xleftarrow{R} \mathbb{Z}_p\right\}\ .$$

The DDH assumption states that the advantage of a polynomial-time distinguisher $\mathcal{A}$, defined as

$$\mathbf{Adv}^{\mathsf{DDH}}(\mathcal{A}) = \Big|\Pr[\mathcal{A}(g,u,v,w) = 1 \mid (g,u,v,w) \xleftarrow{R} D_0] - \Pr[\mathcal{A}(g,u,v,w) = 1 \mid (g,u,v,w) \xleftarrow{R} D_1]\Big|$$

is negligible.

We are now ready to present the scheme. It is given by the following tuple of algorithms.

Setup(1^λ). Let $\mathbb{G}$ be a group of prime order $M = p$ for which the DDH assumption holds, and let $g \in \mathbb{G}$ be a random generator. Let also $H_1 : \mathbb{Z} \to \mathbb{G}$ and $H_2 : \mathbb{Z} \to \mathbb{G}$ be two hash functions. Finally, for $2n$ random elements $s_1, \ldots, s_n, t_1, \ldots, t_n \xleftarrow{R} \mathbb{Z}_p$, define $s_0 = -\sum_{i=1}^n s_i \bmod p$ and $t_0 = -\sum_{i=1}^n t_i \bmod p$.

The system parameters are $\mathsf{param} = \{p, \mathbb{G}, g, H_1, H_2\}$ and the secret key of user i is $\mathsf{sk}_i = (s_i, t_i)$, with $0 \leq i \leq n$.

$\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, \tau, x_{i,\tau})$. At time period τ, for a private input $x_{i,\tau} \in \mathbb{Z}_p$, user i produces

$$c_{i,\tau} = g^{x_{i,\tau}} H_1(\tau)^{s_i} H_2(\tau)^{t_i} \ .$$

$\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, \tau, c_{1,\tau}, \ldots, c_{n,\tau})$. The aggregator obtains the sum $X_\tau = \sum_{i=1}^{n} x_{i,\tau}$ for time period τ by first computing $V_\tau := H_1(\tau)^{s_0} H_2(\tau)^{t_0} \prod_{i=1}^{n} c_{i,\tau} = g^{X_\tau}$ and next the discrete logarithm of V_τ w.r.t. basis g.

As for the Shi *et al.* construction, since g has order p, the sum X_τ is computed modulo $M = p$. Further, in the $\mathsf{AggrDec}$ algorithm, the aggregator has to obtain the value of X_τ from $V_\tau = g^{X_\tau}$ in $\mathbb{G}$. The most appropriate method for computing discrete logarithms is Pollard's λ algorithm (or variants thereof) and requires that the range of X_τ be relatively small.

3.2. Security

The next theorem proves that the basic scheme meets the AO security notion in the random oracle model, based on the DDH assumption.

THEOREM 3.2. *The scheme provides* AO *security under the* DDH *assumption in the random oracle model. Specifically, for any probabilistic polynomial-time adversary* $\mathcal{A}$*, there exists a* DDH *distinguisher* $\mathcal{B}_{\mathsf{DDH}}$ *with comparable running time*[1] *and such that*

$$\mathbf{Adv}^{\mathsf{AO}}(\mathcal{A}) \leq 2e\,(q_{enc} + 1) \cdot (\mathbf{Adv}^{\mathsf{DDH}}(\mathcal{B}_{\mathsf{DDH}}) + 1/p)\,,$$

where q_{enc} *is the number of encryption queries made by the adversary for distinct time periods other than* $\tau^\star$*, and* e *is the base for the natural logarithm.*

PROOF. The theorem is a corollary of Theorem 4.3, which ensures the security of our abstract scheme (cf. Section 4.4). □

3.3. Performance

The new scheme combines the advantages of [Joye and Libert 2013] (which offers tighter security in the random oracle model) and of [Shi et al. 2011] (which has more compact ciphertexts when implemented using elliptic-curve groups). As already shown in Table I (Section 1), our basic scheme represents the aggregator-oblivious encryption with the shortest ciphertexts. The key sizes are derived from the ECRYPT 2 recommendations [ECRYPT II 2012] —where Shi *et al.*'s scheme and our basic scheme are implemented in elliptic-curve groups and Joye-Libert's scheme in $\mathbb{Z}_{N^2}^*$ with N an RSA modulus. Concretely, for a security gap of 2^γ, key sizes are chosen such that the underlying problem (DDH for the Shi *et al.*'s scheme and our scheme, or DCR for the Joye-Libert scheme) cannot be solved in time $2^{\gamma+\lambda}$ (with constant probability, e.g., 1/2). According to ECRYPT 2 recommendations, for DDH-based scheme, the order of the group used has to have $2(\gamma + \lambda)$ bits, while for DCR-based scheme, the modulus has to have $s^{-1}(\gamma + \lambda)$ bits, where s is the function defined in [ECRYPT II 2012, Section 6.2.1]. As the latter function is sublinear, higher security levels yield better results for our basic scheme compared to [Joye and Libert 2013].

As exemplified in Table II, our new scheme also features better encryption times. Table II presents the running times for a security level of 80 bits and was constructed from a synthetically generated dataset by taking $n = 2^{20} \approx 10^6$ users and $T = 2^{20}$ time periods. This

[1] By "comparable running time", we mean that $\mathcal{B}_{\mathsf{DDH}}$'s running time exceeds that of $\mathcal{A}$ by a small *additive* term (rather than a multiplicative factor) which only depends on the security parameter and the number of queries, but not on $\mathcal{A}$'s running time. More precisely, $\mathcal{B}_{\mathsf{DDH}}$'s running time roughly amounts to the running time of $\mathcal{A}$ plus $O(q_H + q_{enc})$ group exponentiations, where q_H is the number of queries to the random oracle.

approximately allows computing an aggregation every 15 minutes for 30 years over a city like Paris (there are about one million households in Paris). Moreover, to have the fairest possible comparison, the worst case for our reduction is considered: $q_{enc} = T - 1 \approx 2^{20}$. Higher security levels yield better results for our basic scheme compared to [Joye and Libert 2013] as attacks against factorization and DCR are subexponential in the key size while attacks against DDH are exponential in the key size.

Table II. Running times (with margin of error at 95% confidence, computed with 100 samples) of our basic scheme and the existing schemes (for $T = n = 2^{20}$, 80-bit security, using parameters in Table I, SHA-512 for hashing, 24-bit $x_{i,\tau}$, and X_τ in a pre-determined 24-bit range, on an Intel™ Core i5 750 with MIRACL™ library https://github.com/CertiVox/MIRACL, Jun 20, 2013)

Scheme	Time (ms)			
	Hashing[a]	Encryption[b]	First phase of decryption[c]	Second phase of decryption[d]
[Shi et al. 2011]	0.23 (±0.01)	5.5 (±0.1)	11.3 (±0.0)	192 (±20)
[Joye and Libert 2013]	0.01 (±0.00)	58.3 (±0.5)	45.5 (±0.0)	0.0 (±0.0)
Our basic scheme	**0.23 (±0.01)**	**2.6 (±0.1)**	**6.9 (±0.0)**	**126 (±13)**

[a] Computation of $H(\tau)$ or $H_1(\tau)$ and $H_2(\tau)$;
[b] Computation of $c_{i,\tau}$, excluding computation of $H(\tau)/H_1(\tau)/H_2(\tau)$;
[c] Computation of V_τ from $(c_{i,\tau})$, excluding computation of $H(\tau)/H_1(\tau)/H_2(\tau)$;
[d] Computation of X_τ from V_τ (we used a variant of the Pollard's kangaroo (or λ) method described in [Montenegro and Tetali 2009]).

4. GENERALIZATION USING KEY-HOMOMORPHIC SMOOTH PROJECTIVE HASH FUNCTIONS

In this section, we use the framework of key-homomorphic smooth projective hashing to generalize our DDH construction presented in Section 3.

4.1. Key-homomorphic smooth projective hash functions

4.1.1. Subset-membership problem. We start with the important notion of subset-membership problems, as introduced in [Cramer and Shoup 2002]. Consider an NP-language $\mathcal{L} \subset \mathcal{X}$, defined by a polynomial-time witness relation $\mathcal{R}$:

$$\mathcal{L} = \left\{ y \in \mathcal{X} \mid \exists w \text{ such that } \mathcal{R}(y, w) = 1 \right\} .$$

We suppose that $\mathcal{L}$, $\mathcal{X}$, $\bar{\mathcal{L}} = \mathcal{X} \setminus \mathcal{L}$ are efficiently (uniformly) samplable, and even that sampling a word $y \in \mathcal{L}$ along with an associated witness w for this word can also be done efficiently. We also assume that $|\mathcal{L}|/|\mathcal{X}|$ is negligible: in other words, a random element of $\mathcal{X}$ is in $\bar{\mathcal{L}}$ with overwhelming probability.

Basically, the language $\mathcal{L}$ induces a hard subset-membership problem if random elements of $\mathcal{L}$ cannot be distinguished from random elements of $\mathcal{X}$. More formally, this notion can be defined via the following game between a challenger and an attacker. The challenger chooses at random a bit b. The attacker can issue up to q_m (a parameter of the game) queries to the challenger. On each query, the challenger returns a uniformly random element in $\mathcal{X}$ if $b = 0$, and a uniformly random element in $\mathcal{L}$ if $b = 1$. At the end of the game, the attacker outputs a bit b' and wins the game if and only if $b' = b$.

Definition 4.1. A subset-membership problem is *hard* if, in the previous game, the advantage, which is defined as $\mathbf{Adv}^{\mathsf{memb}}_{q_m}(\mathcal{A}) := 2\left|\Pr[b' = b] - 1/2\right|$, is negligible for any probabilistic polynomial-time attacker $\mathcal{A}$.

Defining the subset-membership hardness via the above game allows us to generically obtain tighter security bounds. In instantiations based on specific assumptions (e.g., DDH), the random self-reducibility of underlying problems (e.g., [Abadi et al. 1989; Stadler 1996; Naor and Reingold 1997]) allows avoiding any dependency on q_m in the reduction.

4.1.2. Smooth projective hash functions. Smooth projective hash functions (SPHFs) were introduced by Cramer and Shoup [2002] as a tool to construct chosen-ciphertext secure encryption schemes. We present below a variant tailored to fulfill our needs. A recent account on the different flavors of SPHFs can be found in [Benhamouda et al. 2013].

Definition 4.2. Using the previous notations, a *smooth projective hash function* (SPHF) is specified by a tuple of algorithms, $(\mathsf{HashKG}, \mathsf{ProjKG}, \mathsf{Hash}, \mathsf{ProjHash})$, of which the first one is probabilistic and the other algorithms are deterministic, defined as:

$\mathsf{HashKG}(1^\lambda)$. On input a security parameter λ, algorithm HashKG generates a hashing key hk identified as an element of $\mathcal{K}$, where $\mathcal{K}$ denotes the key space.
$\mathsf{ProjKG}(\mathsf{hk})$. Given a hashing key hk, this algorithm derives a projection key hp.
$\mathsf{Hash}(\mathsf{hk}, y)$. Given a hashing key hk and a word $y \in \mathcal{X}$, algorithm Hash outputs the hash value h of y.
$\mathsf{ProjHash}(\mathsf{hp}, y, w)$. Given a projection key hp, a word $y \in \mathcal{L}$ and a corresponding witness w (such that $\mathcal{R}(y, w) = 1$), algorithm ProjHash outputs the hash value h of y.

Further, letting Π denote the range of Hash and ProjHash and assuming that $(\Pi, \cdot)$ is an Abelian group (written multiplicatively with 1_Π as neutral element), an SPHF must satisfy the properties of *correctness* and *special smoothness*:

Correctness. This property means that Hash and ProjHash hash to the same value for any word y in the language $\mathcal{L}$. More precisely, for every hashing key $\mathsf{hk} \xleftarrow{R} \mathsf{HashKG}(1^\lambda)$, for all $y \in \mathcal{L}$ and associated witness w (such that $\mathcal{R}(y, w) = 1$), we have

$$\mathsf{Hash}(\mathsf{hk}, y) = \mathsf{ProjHash}(\mathsf{hp}, y, w)$$

provided $\mathsf{hp} = \mathsf{ProjHash}(\mathsf{hk})$.

Special smoothness. Let $\Pi' \subseteq \Pi$ be a subset of Π. Intuitively, the special smoothness says that the hash value of any $y \in \bar{\mathcal{L}} = \mathcal{X} \setminus \mathcal{L}$ looks random "over Π'", even knowing hp. Formally, an SPHF is said to be (ε_s, Π')-*smooth*, if for all $y \in \bar{\mathcal{L}}$, the following two distributions are ε_s-statistically indistinguishable:

$$D_0 = \left\{(\mathsf{hp}, h) \mid \mathsf{hk} \xleftarrow{R} \mathcal{K},\ \mathsf{hp} \leftarrow \mathsf{ProjKG}(\mathsf{hk}),\ h \leftarrow \mathsf{Hash}(\mathsf{hk}, y)\right\}$$

and

$$D_1 = \left\{(\mathsf{hp}, h \cdot h') \mid \mathsf{hk} \xleftarrow{R} \mathcal{K},\ \mathsf{hp} \leftarrow \mathsf{ProjKG}(\mathsf{hk}), h \leftarrow \mathsf{Hash}(\mathsf{hk}, y),\ h' \xleftarrow{R} \Pi'\right\} .$$

There are a couple of differences compared to the definition given in [Cramer and Shoup 2002]. Special smoothness replaces the original smoothness property. The latter basically corresponds to the definition of special smoothness when $\Pi' = \Pi$. The special smoothness is also required to hold for *any* word $y \in \bar{\mathcal{L}}$, and not just on average as in the original definition. Furthermore, only HashKG and Hash are required to be polynomial-time algorithms; ProjKG or ProjHash are not required to be efficient and they may run in exponential time.

An additional property which will be used in the security proofs is called *key uniformity*.

Key uniformity. An SPHF is $\varepsilon_{\mathsf{hk}}$-*key-uniform* when an honestly generated hashing key is $\varepsilon_{\mathsf{hk}}$-statistically indistinguishable[2] from a random element in $\mathcal{K}$.

[2]Recall that two distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ on some finite set S are ε-*statistically indistinguishable* or ε-*close* if $\frac{1}{2}\sum_{x \in S} |\Pr_{X \xleftarrow{R} \mathcal{D}_0}[X = x] - \Pr_{X \xleftarrow{R} \mathcal{D}_1}[X = x]| \leq \varepsilon$.

4.1.3. Key-homomorphic SPHF. An SPHF is said *key-homomorphic* if in addition

(1) $(\mathcal{K}, +)$ is an Abelian group (written additively with $0_{\mathcal{K}}$ as neutral element);
(2) Π' is a subgroup of Π;
(3) for any two hashing keys $\mathsf{hk}_1, \mathsf{hk}_2 \in \mathcal{K}$ and for every word $y \in \mathcal{X}$

$$\mathsf{Hash}(\mathsf{hk}_1 + \mathsf{hk}_2, y) = \mathsf{Hash}(\mathsf{hk}_1, y) \cdot \mathsf{Hash}(\mathsf{hk}_2, y) \ .$$

4.2. Our Abstract Scheme

We can now present our generic aggregator-oblivious encryption scheme, based on key-uniform, key-homomorphic SPHF with the special smoothness property.

Let f be an injective group homomorphism, $f : \mathbb{Z}_M \to \Pi'$. We assume that this homomorphism is efficiently and publicly invertible over the domain of possible sums X_τ (which may be smaller than $\{0, \ldots, M-1\}$, as is the case for our DDH-based scheme).

$\mathsf{Setup}(1^\lambda)$. Let $\mathcal{L} \subset \mathcal{X}$ be a hard subset-membership language. Let also $H : \mathbb{Z} \to \mathcal{X}$ be a hash function (viewed as a random oracle in the security analysis). Finally, let $\mathsf{hk}_1 \xleftarrow{R} \mathsf{HashKG}(1^\lambda), \ldots, \mathsf{hk}_n \xleftarrow{R} \mathsf{HashKG}(1^\lambda)$ be n random hashing key. Define the aggregator's secret key as $\mathsf{hk}_0 = -\sum_{i=1}^n \mathsf{hk}_i$.

The system parameters are $\mathsf{param} = \{\mathcal{L}, H\}$ and the secret key of user i is $\mathsf{sk}_i = \mathsf{hk}_i$, with $0 \le i \le n$.

$\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, \tau, x_{i,\tau})$. At time period τ, for a private input $x_{i,\tau} \in \mathbb{Z}_M$, user i produces

$$c_{i,\tau} = f(x_{i,\tau}) \cdot \mathsf{Hash}(\mathsf{hk}_i, H(\tau)) \in \Pi \ .$$

$\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, \tau, c_{1,\tau}, \ldots, c_{n,\tau})$. The aggregator obtains the sum $X_\tau \pmod M$ for time period τ by computing $X_\tau := f^{-1}\big(\mathsf{Hash}(\mathsf{hk}_0, H(\tau)) \cdot \prod_{i=1}^n c_{i,\tau}\big)$.

The correctness of the aggregation step follows from the homomorphic properties:

$$\begin{aligned}
\mathsf{Hash}(\mathsf{hk}_0, H(\tau)) \cdot \textstyle\prod_{i=1}^n c_{i,\tau} &= \textstyle\prod_{i=0}^n \mathsf{Hash}(\mathsf{hk}_i, H(\tau)) \cdot \prod_{i=1}^n f(x_{i,\tau}) \\
&= \mathsf{Hash}(\textstyle\sum_{i=0}^n \mathsf{hk}_i, H(\tau)) \cdot f(\sum_{i=1}^n x_{i,\tau}) \\
&= \mathsf{Hash}(0_{\mathcal{K}}, H(\tau)) \cdot f(\textstyle\sum_{i=1}^n x_{i,\tau}) = 1_\Pi \cdot f(\sum_{i=1}^n x_{i,\tau}) \\
&= f(\textstyle\sum_{i=1}^n x_{i,\tau})
\end{aligned}$$

and therefore $X_\tau = \sum_{i=1}^n x_{i,\tau} \pmod M$ since f is injective.

4.3. Security

We prove security of the abstract scheme in the random oracle model, based on the hardness of the subset-membership problem.

THEOREM 4.3. *The scheme provides* AO *security under the hard subset-membership assumption of $\mathcal{L}$ in the random oracle model. Namely, for any probabilistic polynomial-time adversary $\mathcal{A}$, if the SPHF is (ε_s, Π')-smooth and $\varepsilon_{\mathsf{hk}}$-key-uniform, there exists a distinguisher $\mathcal{B}_{\mathsf{memb}}$ for the subset-membership problem of $\mathcal{L}$ with comparable running time and such that*

$$\mathbf{Adv}^{\mathsf{AO}}(\mathcal{A}) \le 2e\,(q_{enc}+1) \cdot \left(\mathbf{Adv}_T^{\mathsf{memb}}(\mathcal{B}_{\mathsf{memb}}) + n\,\varepsilon_{\mathsf{hk}} + \frac{|\mathcal{L}|}{|\mathcal{X}|} + n(n+1)\,\varepsilon_s \right) ,$$

where n is the number of users, q_{enc} is the number of encryption queries made by the adversary for distinct time periods other than $\tau^\star$, and e is the base for the natural logarithm.

We recall that we suppose $|\mathcal{L}|/|\mathcal{X}|$ is negligible.

PROOF. The proof proceeds with a sequence of several games. It begins with Game 0, which is the real game, and ends with Game 6, where even a computationally unbounded adversary has no advantage. For each $j \in \{0, \ldots, 6\}$, we denote by S_j the event that the challenger $\mathcal{B}$ outputs 1 in Game j.[3] We also define $Adv_j = 2 \cdot |\Pr[S_j] - 1/2|$.

In the following, we assume w.l.o.g. that the adversary $\mathcal{A}$ has always already queried the random oracle H on input τ before any encryption query for the time period τ. We assume that the adversary does not query the random oracle H more than once for a given τ.

Game 0. This is the real game. Namely, the challenger performs the setup of the system by generating $\mathsf{hk}_1, \ldots, \mathsf{hk}_n$ using HashKG and defining $\mathsf{hk}_0 = -\sum_{i=1}^{n} \mathsf{hk}_i$. Queries to the random oracle H are answered by returning uniformly random group elements in $\mathcal{X}$. Encryption queries $(i, \tau, x_{i,\tau})$ for time period τ are answered by returning the ciphertext $c_{i,\tau} = f(x_{i,\tau}) \cdot h_{i,\tau}$ with $h_{i,\tau} = \mathsf{Hash}(\mathsf{hk}_i, H(\tau))$. Whenever the adversary decides to corrupt some player $i \in \{0, \ldots, n\}$, the challenger reveals hk_i. In the challenge phase, the adversary chooses a target time period $\tau^\star$, an uncorrupted subset $\mathcal{S}^\star \subseteq \mathcal{U}^\star$ and two distinct series $\langle (i, \tau^\star, x^{(0)}_{i,\tau^\star}) \rangle_{i \in \mathcal{S}^\star}$, $\langle (i, \tau^\star, x^{(1)}_{i,\tau^\star}) \rangle_{i \in \mathcal{S}^\star}$ which must sum up to the same value if $\mathcal{S}^\star = \mathcal{U}^\star$ and the aggregator's private key sk_0 is exposed at some point of the game (see Section 2.1). At this stage, the challenger flips a fair binary coin $b \xleftarrow{R} \{0,1\}$ and the adversary $\mathcal{A}$ receives

$$\left\{ c_{i,\tau^\star} = f(x^{(b)}_{i,\tau^\star}) \cdot h_{i,\tau^\star} \right\}_{i \in \mathcal{S}^\star} \quad \text{with } h_{i,\tau^\star} = \mathsf{Hash}(\mathsf{hk}_i, H(\tau^\star)) \ .$$

We assume that the adversary queries $H(\tau^\star)$ before the challenge phase. Otherwise, $\mathcal{B}$ can simply make the query for itself. In the second phase, after a second series of queries, $\mathcal{A}$ outputs a bit $b' \in \{0,1\}$. We let the challenger $\mathcal{B}$ output 1 if $b' = b$ and 0 otherwise. The adversary's advantage in Game 0 is thus $Adv_0 = 2 \cdot |\Pr[S_0] - 1/2| = \mathbf{Adv}^{\mathsf{AO}}(\mathcal{A})$.

Game 1. This game is identical to Game 0 with the following difference. For each random oracle query $H(\tau)$, the challenger $\mathcal{B}$ flips a biased coin $\delta_\tau \in \{0,1\}$ that takes the value 1 with probability $1/(q_{enc}+1)$ and the value 0 with probability $q_{enc}/(q_{enc}+1)$. At the end of the game, $\mathcal{B}$ considers the event E that one of the following conditions holds:

— For the target time period $\tau^\star$, the coin $\delta_{\tau^\star}$ flipped for the hash query $H(\tau^\star)$ was $\delta_{\tau^\star} = 0$.
— There exists a time period $\tau \neq \tau^\star$ such that an encryption query $(i, \tau, .)$ was made for some user $i \in \mathcal{U}^\star$ but for which $\delta_\tau = 1$.

If event E occurs (which $\mathcal{B}$ can detect at the end of the game), $\mathcal{B}$ halts and outputs a random bit. Otherwise, it outputs 1 if and only if $b' = b$. An analysis similar to that in [Coron 2000] shows that

$$\Pr[\neg E] = \frac{1}{q_{enc}+1} \cdot \left(\frac{q_{enc}}{q_{enc}+1} \right)^{q_{enc}} = \frac{1}{q_{enc}+1} \left(1 - \frac{1}{q_{enc}+1} \right)^{q_{enc}} \geq \frac{1}{e\,(q_{enc}+1)} \,,$$

where e is the base for the natural logarithm. The transition from Game 0 to Game 1 is thus a transition based on a failure event of large probability [Dent 2006] and we therefore have $Adv_1 = Adv_0 \cdot \Pr[\neg E] \geq Adv_0/(e(q_{enc}+1))$.

[3] In the proof, the output $b' \in \{0,1\}$ of the adversary $\mathcal{A}$ is not directly the output of the game, but is first given to the challenger which then outputs a bit. For Game 0, which corresponds to the AO security notion, this modification does not change anything. But it helps for the other games, as it enables the challenger to change the output of the adversary.

Game 2. In this game, we modify the distribution of random oracle outputs. Specifically, the treatment of each hash query τ depends on the random coin $\delta_\tau \in \{0,1\}$.

— If $\delta_\tau = 0$, the challenger $\mathcal{B}$ samples a random word $y_\tau \in \mathcal{L}$ together with a witness w_τ, and defines $H(\tau) = y_\tau$;
— If $\delta_\tau = 1$, $\mathcal{B}$ samples a word $y_\tau \in \mathcal{X}$ and defines $H(\tau) = y_\tau$.

It is straightforward to see that Game 2 and Game 1 are computationally indistinguishable if the hard subset-membership assumption holds. Namely, there exists a distinguisher $\mathcal{B}_{\mathsf{memb}}$ for the subset-membership problem of $\mathcal{L}$ with comparable running time and such that

$$|\Pr[S_2] - \Pr[S_1]| \leq \mathbf{Adv}_T^{\mathsf{memb}}(\mathcal{B}_{\mathsf{memb}}) \ .$$

Therefore $Adv_1 \leq Adv_2 + 2 \cdot \mathbf{Adv}_T^{\mathsf{memb}}(\mathcal{B}_{\mathsf{memb}})$.

Game 3. In this game, after the challenge query, if $H(\tau^\star) = y_{\tau^\star} \in \mathcal{L}$, then $\mathcal{B}$ halts and outputs a random bit. We recall that, in the previous game, $y_{\tau^\star}$ was drawn uniformly at random from $\mathcal{X}$ (or $\mathcal{B}$ would already have stopped because of event E). Therefore $|\Pr[S_3] - \Pr[S_2]| \leq |\mathcal{L}|/|\mathcal{X}|$ and $Adv_2 \leq Adv_3 + 2|\mathcal{L}|/|\mathcal{X}|$.

Game 4. In this game, $\mathcal{B}$ generates hk_i as $\mathsf{hk}_i \xleftarrow{R} \mathcal{K}$, for each $i \in \{1, \ldots, n\}$, instead of using HashKG (note that $\mathcal{B}$ is not polynomially bounded in this game nor in the subsequent games). Therefore $|\Pr[S_4] - \Pr[S_3]| \leq n\,\varepsilon_{\mathsf{hk}}$ and $Adv_3 \leq Adv_4 + 2n\,\varepsilon_{\mathsf{hk}}$.

Game 5. At the beginning of the game, the challenger picks a random index $i^\star \in \{0, \ldots, n\}$ and at the end of the game, $\mathcal{B}$ considers event E' that one of the following conditions holds:

— $i^\star \neq \max \mathcal{S}^\star$ and $\sum_{i \in \mathcal{S}^\star} x_{i,\tau^\star}^{(0)} = \sum_{i \in \mathcal{S}^\star} x_{i,\tau^\star}^{(1)}$; or
— $i^\star \neq \max(\mathcal{U}'^\star \setminus \mathcal{S}^\star)$ and $\sum_{i \in \mathcal{S}^\star} x_{i,\tau^\star}^{(0)} \neq \sum_{i \in \mathcal{S}^\star} x_{i,\tau^\star}^{(1)}$, where

$$\mathcal{U}'^\star = \begin{cases} \mathcal{U}^\star & \text{when the aggregator is corrupted}\,, \\ \mathcal{U}^\star \cup \{0\} & \text{otherwise}\,. \end{cases}$$

Notice that when $\sum_{i \in \mathcal{S}^\star} x_{i,\tau^\star}^{(0)} \neq \sum_{i \in \mathcal{S}^\star} x_{i,\tau^\star}^{(1)}$, the set $\mathcal{U}'^\star \setminus \mathcal{S}^\star$ cannot be empty, so that the previous definition makes sense. If event E occurs (which $\mathcal{B}$ can detect at the end of the game), $\mathcal{B}$ halts and outputs a random bit. Otherwise, it outputs 1 if and only if $b' = b$. As for the transition from Game 0 to Game 1, the transition from Game 4 to Game 5 is a transition based on a failure event of large probability. We therefore have $Adv_5 = Adv_4 \cdot \Pr[\neg E'] = Adv_4/(n+1)$.

Game 6. In this game, $\mathcal{B}$ picks hk_i independently and uniformly at random in $\mathcal{K}$, for $i \in \{0, \ldots, n\} \setminus \{i^\star\}$, and computes $\mathsf{hk}_{i^\star} = -\sum_{i \in \{0,\ldots,n\} \setminus \{i^\star\}} \mathsf{hk}_i$ at the beginning of the game. For any time period τ, $\mathcal{B}$ computes $h_{i^\star,\tau}$ as $\prod_{i \in \{0,\ldots,n\} \setminus \{i^\star\}} (h_{i,\tau})^{-1}$. Furthermore, for any $i \in \mathcal{U}'^\star \setminus \{i^\star\}$ and any time period $\tau \neq \tau^\star$, $\mathcal{B}$ now computes $h_{i,\tau}$ as $\mathsf{ProjHash}(\mathsf{hp}_i, y_\tau, w_\tau)$ (with w_τ a witness for $y_\tau \in \mathcal{L}$), instead of $\mathsf{Hash}(\mathsf{hk}_i, y_\tau)$. Game 6 is perfectly indistinguishable from Game 5 (thanks to the fact that $\mathsf{hk}_0, \ldots, \mathsf{hk}_n$ are, in both games, all chosen uniformly at random under the only condition that $\sum_{i=0}^n \mathsf{hk}_i = 0$, and thanks to key-homomorphism and correctness of the SPHF) and $Adv_6 = Adv_5$.

Game 7. In this game, for all $i \in \mathcal{U}^\star \setminus \{i^\star\}$, we set $h_{i,\tau^\star} = \mathsf{Hash}(\mathsf{hk}_i, H(\tau^\star)) \cdot h'_i$ with $h'_i \xleftarrow{R} \Pi'$. In other words, we have:

— if $\sum_{i\in\mathcal{S}^\star} x^{(0)}_{i,\tau^\star} \neq \sum_{i\in\mathcal{S}^\star} x^{(1)}_{i,\tau^\star}$, then $h_{i,\tau^\star} = \mathsf{Hash}(\mathsf{hk}_i, H(\tau^\star)) \cdot h'_i$ with $h'_i \xleftarrow{R} \Pi'$, for $i \in \mathcal{S}^\star \subseteq \mathcal{U}^\star \setminus \{i^\star\}$; in this case, clearly, h'_i completely masks $f(x^{(b)}_{i,\tau^\star})$;

— otherwise, then $h_{i,\tau^\star} = \mathsf{Hash}(\mathsf{hk}_i, H(\tau^\star))\cdot h'_i$ with $h'_i \xleftarrow{R} \Pi'$ for $i \in \mathcal{S}^\star\setminus\{i^\star\} \subseteq \mathcal{U}^\star\setminus\{i^\star\}$; and we have that, for any sequence $(h'_i)_{i\in\mathcal{S}^\star}$, there exists a unique sequence $(h''_i)_{i\in\mathcal{S}^\star}$ (with $\prod_{i\in\mathcal{S}^\star} h'_i = \prod_{i\in\mathcal{S}^\star} h''_i = 1$), that satisfies $h'_i \cdot f(x^{(0)}_{i,\tau^\star}) = h''_i \cdot f(x^{(1)}_{i,\tau^\star})$.

It is therefore clear that $\Pr[S_7] = 1/2$, so that $\mathcal{A}$ has no advantage ($Adv_7 = 0$).

In addition, Games 6 and 7 are statistically indistinguishable thanks to the special smoothness of the SPHF: $|\Pr[S_7] - \Pr[S_6]| \leq n\,\varepsilon_s$ and so $Adv_6 \leq Adv_7 + 2n\,\varepsilon_s$.

Putting all together, we get

$$\mathbf{Adv}^{\mathsf{AO}}(\mathcal{A}) \leq 2e\,(q_{enc}+1)\cdot\left(\mathbf{Adv}^{\mathsf{memb}}(\mathcal{B}) + n\,\varepsilon_{\mathsf{hk}} + \tfrac{|\mathcal{L}|}{|\mathcal{X}|} + n(n+1)\,\varepsilon_s\right),$$

which concludes the proof. □

4.4. A concrete instantiation

An example of hard subset-membership language is the DDH language. We then have:

$$\mathcal{X} = \mathbb{G}\times\mathbb{G}, \quad \mathcal{L} = \left\{\vec{y} = (y_1, y_2) \in \mathcal{X} \mid \exists w \in \mathbb{Z}_p \text{ such that } \mathcal{R}(\vec{y}, w) = 1\right\}$$

with

$$\mathcal{R}(\vec{y}, w) = 1 \iff y_1 = {g_1}^w \text{ and } y_2 = {g_2}^w,$$

where g_1 and g_2 are two generators of a group $\mathbb{G}$ of prime order p. The hard subset-membership assumption for this language is the DDH assumption.

For the DDH language, Cramer and Shoup construct an SPHF as follows. The key space is $\mathcal{K} = \mathbb{Z}_p\times\mathbb{Z}_p$ and the hash range is $\Pi = \mathbb{G}$. The hashing key is a random tuple $\mathsf{hk} = (s,t) \in \mathcal{K}$ and the projection key is $\mathsf{hp} = {g_1}^s {g_2}^t$. We then have:

$$\mathsf{Hash}\colon \mathcal{K}\times\mathcal{X}\to\Pi,\ \big(\mathsf{hk}, (y_1,y_2)\big) \mapsto \mathsf{Hash}\big(\mathsf{hk}, (y_1,y_2)\big) = {y_1}^s {y_2}^t$$

and, if w is a witness for (y_1, y_2), i.e., $y_1 = {g_1}^w$ and $y_2 = {g_2}^w$,

$$\mathsf{ProjHash}\colon \mathbb{G}\times\mathcal{X}\times\mathbb{Z}_p\to\Pi,\ \big(\mathsf{hp}, (y_1,y_2), w\big) \mapsto \mathsf{ProjHash}\big(\mathsf{hp}, (y_1,y_2), w\big) = \mathsf{hp}^w\ .$$

This SPHF is $(0,\Pi)$-smooth (see [Cramer and Shoup 2002]). It is readily seen that this SPHF is key-homomorphic and 0-key-uniform. If we set $f : \mathbb{Z}_p \to \Pi, x \mapsto f(x) = g^x$ (where g is a generator of $\mathbb{G}$), which clearly is an injective group homomorphism, we get exactly our DDH-based scheme of Section 3. Observe also that plugging $\varepsilon_{\mathsf{hk}} = \varepsilon_s = 0$, $|\mathcal{X}| = p^2$, and $|\mathcal{L}| = p$ in Theorem 4.3 yields the statement of Theorem 3.2.

5. FURTHER INSTANTIATIONS

5.1. k-linear assumption and generalizations

In [2013], Escala *et al.* generalize the k-LIN assumptions [Boneh et al. 2004; Hofheinz and Kiltz 2007; Shacham 2007] in an assumption called MDDH. They also show how to construct an SPHF from any MDDH assumption. Their construction can be used directly in our framework, with $f(x) = g^x$ (which is invertible for x in small ranges), since their SPHFs are clearly key-homomorphic and 0-key-uniform. This yields in particular an aggregator oblivious encryption scheme from the k-LIN assumption.

When $k = 1$, since 1-LIN = DDH, we get exactly the new scheme presented in Section 3. A larger value of k implies a weaker assumption: indeed, the k-LIN assumption is implied by the $(k-1)$-LIN assumption for any $k > 1$, as shown in [Hofheinz and Kiltz 2007; Shacham 2007]. The disadvantage of a larger k is an increase of the private-key size, which has to be

multiplied by $(k+1)/2$. However, other parameter sizes given in Table I remain unchanged. In particular, increasing k does not affect the ciphertext size, which is unusual for encryption schemes based on the k-LIN assumption.

5.2. SD assumption

Yet another instantiation can be derived from the subgroup decision (SD) assumption (for cyclic groups of composite order, without pairing) as introduced by Boneh et al. [2005].

Let $N = pq$ be an RSA modulus with p and q two large primes. Let $\mathbb{G} = \langle g \rangle$ be a cyclic group of order N, $\mathbb{G}_p = \langle g_p \rangle$ be the subgroup of order p, and $\mathbb{G}_q = \langle g_q \rangle$ be the subgroup of order q. Define $\mathcal{X} = \mathbb{G}$, $\mathcal{L} = \mathbb{G}_p$, $\mathcal{K} = \mathbb{Z}_N$, $\Pi = \mathbb{G}$, and $\Pi' = \mathbb{G}_q$. A witness w for $y \in \mathcal{L}$ is a discrete logarithm of y in base g_p, $y = {g_p}^w$. Then set:

— for any $y \in \mathcal{X}$, $\mathsf{Hash}(\mathsf{hk}, y) = y^r$, and for any $y \in \mathcal{L}$, $\mathsf{ProjHash}(\mathsf{hp}, y, w) = \mathsf{hp}^w$ with w a witness for y such that $y = {g_p}^w$, for $\mathsf{hk} = r \xleftarrow{R} \mathcal{K}$ and $\mathsf{hp} = {g_p}^{\mathsf{hk}}$;
— $f(x) = {g_q}^x$ for $x \in \mathbb{Z}_q$.

The so-obtained SPHF is $(0, \Pi')$-smooth, 0-key-uniform, and key-homomorphic.

5.3. DCR assumption

Paillier's decision composite residuosity (DCR) assumption [Paillier 1999] can also be used to instantiate a slight variant of our general framework. The resulting aggregator-oblivious encryption scheme shares many similarities with the Joye-Libert's construction in [Joye and Libert 2013] but it is not strictly the same scheme.

We rely on a variant of the hash proof system proposed by Cramer and Shoup [2002],with inefficient ProjKG and $\mathsf{ProjHash}$. Let $N = pq$ be an RSA modulus where p and q are two distinct primes. Define

$$\mathcal{X} = (\mathbb{Z}_{N^2})^*\,, \quad \mathcal{L} = \{y = z^N \mid z \in \mathcal{X}\} \subset \mathcal{X}\,, \quad \mathcal{K} = \mathbb{Z}_{N\phi(N)}\,,$$
$$\Pi = \mathcal{X}\,, \quad \Pi' = \langle 1 + N \rangle \subset \Pi\,,$$

and set

— For any $y \in \mathcal{X}$, $\mathsf{Hash}(\mathsf{hk}, y) = y^r$, and for any $y \in \mathcal{L}$, $\mathsf{ProjHash}(\mathsf{hp}, y, \perp) = y^{\mathsf{hp}}$, where the hashing key is chosen as $\mathsf{hk} = r \xleftarrow{R} \{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$ (with κ depending on the expected security —see below) and $\mathsf{hp} = r \bmod \phi(N)$ —since $\phi(N)$ is unknown (otherwise the language is not hard subset membership), note that hp cannot be efficiently computed;
— $f(x) = (1+N)^x \in \Pi'$ for $x \in \mathbb{Z}_N$.

If we set $\mathcal{K} = \mathbb{Z}_{N\phi(N)}$ and identify hashing keys $\mathsf{hk} \in \{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$ as elements of $\mathcal{K}$, the resulting SPHF is clearly key-homomorphic. Furthermore, $r \bmod N\phi(N)$ is $1/2^{\kappa+1}$-statistically close from the uniform distribution over $\mathcal{K}$. Since if $r \xleftarrow{R} \mathcal{K}$ would lead to a $(0, \Pi')$-smooth SPHF, the actual resulting SPHF is $(1/2^{\kappa+1}, \Pi')$-smooth. As shown in Appendix B, the corresponding scheme provably meets the notion of aggregator obliviousness. We have the following theorem:

THEOREM 5.1. *The scheme provides* AO *security under the* DCR *assumption in the random oracle model. Specifically, for any probabilistic polynomial-time adversary $\mathcal{A}$, there exists a* DCR *distinguisher $\mathcal{B}_{\mathsf{DCR}}$ with comparable running time and such that*

$$\mathbf{Adv}^{\mathsf{AO}}(\mathcal{A}) \leq 2e\,(q_{enc}+1) \cdot \left(\mathbf{Adv}^{\mathsf{DCR}}(\mathcal{B}) + \frac{n(n+1)}{2^{\kappa+1}} + \frac{1}{\phi(N)} \right)\,,$$

where n is the number of users, q_{enc} is the number of encryption queries made by the adversary for distinct time periods other than $\tau^\star$, and e is the base for the natural logarithm.

REFERENCES

Martín Abadi, Joan Feigenbaum, and Joe Kilian. 1989. On Hiding Information from an Oracle. *J. Comput. System Sci.* 39, 1 (1989), 21–50.

Gergely Ács and Claude Castelluccia. 2011. I Have a DREAM! (DiffeRentially privatE smArt Metering). In *Information Hiding (IH 2011) (LNCS)*, Tomás Filler et al. (Eds.), Vol. 6958. Springer, 118–132.

Mihir Bellare. 1998. Practice-Oriented Provable Security. In *Information Security (ISW '97) (LNCS)*, Eiji Okamoto, George I. Davida, and Masahiro Mambo (Eds.), Vol. 1396. Springer, 221–231.

Mihir Bellare and Phillip Rogaway. 1996. The Exact Security of Digital Signatures: How to Sign with RSA and Rabin. In *EUROCRYPT'96 (LNCS)*, Ueli M. Maurer (Ed.), Vol. 1070. Springer, 399–416.

Fabrice Benhamouda, Olivier Blazy, Céline Chevalier, David Pointcheval, and Damien Vergnaud. 2013. New Techniques for SPHFs and Efficient One-Round PAKE Protocols. In *CRYPTO 2013, Part I (LNCS)*, Ran Canetti and Juan A. Garay (Eds.), Vol. 8042. Springer, 449–475. DOI:http://dx.doi.org/10.1007/978-3-642-40041-4_25

Dan Boneh, Xavier Boyen, and Hovav Shacham. 2004. Short Group Signatures. In *CRYPTO 2004 (LNCS)*, Matthew Franklin (Ed.), Vol. 3152. Springer, 41–55.

Dan Boneh, Eu-Jin Goh, and Kobbi Nissim. 2005. Evaluating 2-DNF Formulas on Ciphertexts. In *TCC 2005 (LNCS)*, Joe Kilian (Ed.), Vol. 3378. Springer, 325–341.

T.-H. Hubert Chan, Elaine Shi, and Dawn Song. 2012. Privacy-Preserving Stream Aggregation with Fault Tolerance. In *FC 2012 (LNCS)*, Angelos D. Keromytis (Ed.), Vol. 7397. Springer, 200–214.

Jean-Sébastien Coron. 2000. On the Exact Security of Full Domain Hash. In *CRYPTO 2000 (LNCS)*, Mihir Bellare (Ed.), Vol. 1880. Springer, 229–235.

Jean-Sébastien Coron. 2002. Optimal Security Proofs for PSS and Other Signature Schemes. In *EUROCRYPT 2002 (LNCS)*, Lars R. Knudsen (Ed.), Vol. 2332. Springer, 272–287.

Ronald Cramer and Victor Shoup. 2002. Universal Hash Proofs and a Paradigm for Adaptive Chosen Ciphertext Secure Public-Key Encryption. In *EUROCRYPT 2002 (LNCS)*, Lars R. Knudsen (Ed.), Vol. 2332. Springer, 45–64.

Alexander W. Dent. 2006. A Note On Game-Hopping Proofs. Cryptology ePrint Archive, Report 2006/260. (2006). http://eprint.iacr.org/.

Cynthia Dwork. 2008. Differential privacy: A survey of results. In *Theory and applications of models of computation*. Springer, 1–19.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *EUROCRYPT 2006 (LNCS)*, Serge Vaudenay (Ed.), Vol. 4004. Springer, 486–503.

ECRYPT II. 2012. Yearly Report on Algorithms and Keysizes. (2012).

Zekeriya Erkin and Gene Tsudik. 2012. Private Computation of Spatial and Temporal Power Consumption with Smart Meters. In *ACNS 12 (LNCS)*, Feng Bao, Pierangela Samarati, and Jianying Zhou (Eds.), Vol. 7341. Springer, 561–577.

Alex Escala, Gottfried Herold, Eike Kiltz, Carla Ràfols, and Jorge Villar. 2013. An Algebraic Framework for Diffie-Hellman Assumptions. In *CRYPTO 2013, Part II (LNCS)*, Ran Canetti and Juan A. Garay (Eds.), Vol. 8043. Springer, 129–147. DOI:http://dx.doi.org/10.1007/978-3-642-40084-1_8

Flavio D. Garcia and Bart Jacobs. 2010. Privacy-Friendly Energy-Metering via Homomorphic Encryption. In *Security and Trust Management (STM 2010) (LNCS)*, Jorge Cuéllar et al. (Eds.), Vol. 6710. Springer, 226–238.

Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. 2013. Candidate Indistinguishability Obfuscation and Functional Encryption for all Circuits. In *54th FOCS*. IEEE Computer Society Press, 40–49.

Shafi Goldwasser, S. Dov Gordon, Vipul Goyal, Abhishek Jain, Jonathan Katz, Feng-Hao Liu, Amit Sahai, Elaine Shi, and Hong-Sheng Zhou. 2014. Multi-input Functional Encryption. In *EUROCRYPT 2014 (LNCS)*, Phong Q. Nguyen and Elisabeth Oswald (Eds.), Vol. 8441. Springer, 578–602. DOI:http://dx.doi.org/10.1007/978-3-642-55220-5_32

Dennis Hofheinz and Eike Kiltz. 2007. Secure Hybrid Encryption from Weakened Key Encapsulation. In *CRYPTO 2007 (LNCS)*, Alfred Menezes (Ed.), Vol. 4622. Springer, 553–571.

Marek Jawurek and Florian Kerschbaum. 2012. Fault-tolerant privacy-preserving statistics. In *Privacy Enhancing Technologies (PETS 2012) (LNCS)*, Simone Fischer-Hübner and Matthew Wright (Eds.), Vol. 7384. Springer, 221–238.

Marek Jawurek, Florian Kerschbaum, and George Danezis. 2012. *Privacy Technologies for Smart Grids – A Survey of Options*. Technical Report MSR-TR-2012-119. Microsoft Research.

Marc Joye and Benoît Libert. 2013. A Scalable Scheme for Privacy-Preserving Aggregation of Time-Series Data. In *FC 2013 (LNCS)*, Ahmad-Reza Sadeghi (Ed.), Vol. 7859. Springer, 111–125. DOI:http://dx.doi.org/10.1007/978-3-642-39884-1_10

Klaus Kursawe, George Danezis, and Markulf Kohlweiss. 2011. Privacy-Friendly Aggregation for the Smart-Grid. In *Privacy Enhancing Technologies (PETS 2011) (LNCS)*, Simone Fischer-Hübner and Nicholas Hopper (Eds.), Vol. 6794. Springer, 221–238.

Iraklis Leontiadis, Kaoutar Elkhiyaoui, and Refik Molva. 2014. Private and Dynamic Time-Series Data Aggregation with Trust Relaxation. In *CANS 14 (LNCS)*, Dimitris Gritzalis, Aggelos Kiayias, and Ioannis G. Askoxylakis (Eds.), Vol. 8813. Springer, 305–320. DOI:http://dx.doi.org/10.1007/978-3-319-12280-9_20

Patrick McDaniel and Stephen McLaughlin. 2009. Security and Privacy Challenges in the Smart Grid. *IEEE Security & Privacy* 7, 3 (May/June 2009), 75–77.

Ravi Montenegro and Prasad Tetali. 2009. How long does it take to catch a wild kangaroo?. In *41st ACM STOC*, Michael Mitzenmacher (Ed.). ACM Press, 553–560.

Moni Naor and Omer Reingold. 1997. Number-theoretic Constructions of Efficient Pseudo-random Functions. In *38th FOCS*. IEEE Computer Society Press, 458–467.

Pascal Paillier. 1999. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In *EUROCRYPT'99 (LNCS)*, Jacques Stern (Ed.), Vol. 1592. Springer, 223–238.

Vibhor Rastogi and Suman Nath. 2010. Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption. In *2010 ACM SIGMOD International Conference on Management of Data (SIGMOD 2010)*, Ahmed K. Elmagarmid and Divyakant Agrawal (Eds.). ACM Press, 735–746.

Hovav Shacham. 2007. A Cramer-Shoup Encryption Scheme from the Linear Assumption and from Progressively Weaker Linear Variants. Cryptology ePrint Archive, Report 2007/074. (2007). http://eprint.iacr.org/.

Elaine Shi, T.-H. Hubert Chan, Eleanor G. Rieffel, Richard Chow, and Dawn Song. 2011. Privacy-Preserving Aggregation of Time-Series Data. In *NDSS 2011*. The Internet Society.

Markus Stadler. 1996. Publicly Verifiable Secret Sharing. In *EUROCRYPT'96 (LNCS)*, Ueli M. Maurer (Ed.), Vol. 1070. Springer, 190–199.

A. AO ENCRYPTION SCHEMES

For completeness, we review in this appendix the two known constructions for aggregator-oblivious encryption [Shi et al. 2011; Joye and Libert 2013].

A.1. Shi *et al.*'s scheme

$\mathsf{Setup}(1^\lambda)$. Let $\mathbb{G}$ be a group of prime order p for which the DDH assumption holds, and let a random generator $g \in \mathbb{G}$. Let also $H : \mathbb{Z} \to \mathbb{G}$ be a hash function viewed as a random oracle. Finally, let n random elements in $\mathbb{Z}_p$, $s_1, \ldots, s_n$, and define $s_0 = -\sum_{i=1}^{n} s_i \bmod p$.

The system parameters are $\mathsf{param} = \{\mathbb{G}, g, H\}$ and the secret key of user i, $0 \leq i \leq n$, is $\mathsf{sk}_i = s_i$.

$\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, \tau, x_{i,\tau})$. At time period τ, for a private input $x_{i,\tau} \in \mathbb{Z}_p$, user i produces

$$c_{i,\tau} = g^{x_{i,\tau}} H(\tau)^{s_i} \ .$$

$\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, \tau, c_{1,\tau}, \ldots, c_{n,\tau})$. The aggregator obtains the sum X_τ for time period τ by first computing $V_\tau := H(\tau)^{s_0} \prod_{i=1}^{n} c_{i,\tau} = g^{X_\tau}$ and next the discrete logarithm of V_τ w.r.t. basis g.

[Notice that since g has order p, the so-obtained value for X_τ is defined modulo $M = p$.]

A.2. Joye-Libert's scheme

$\mathsf{Setup}(1^\lambda)$. Let $M = N = pq$ be an RSA modulus of length ℓ, i.e., a product of two random equal-size primes p, q. Note that the size condition on p and q implies that $\gcd(\phi(N), N) = 1$. Let also $H : \mathbb{Z} \to \mathbb{Z}_{N^2}^*$ be a hash function viewed as a random oracle. Finally let $s_1, \ldots, s_n$ be n randomly chosen elements in $\pm\{0,1\}^{2\ell}$, and define $s_0 = -\sum_{i=1}^{n} s_i$.

The system parameters are $\mathsf{param} = \{N, H\}$ and the secret key of user i, $0 \leq i \leq n$, is $\mathsf{sk}_i = s_i$.

$\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, \tau, x_{i,\tau})$. At time period τ, for a private input $x_{i,\tau} \in \mathbb{Z}_N$, user i produces

$$c_{i,\tau} = (1 + x_{i,\tau}N) \cdot H(\tau)^{s_i} \bmod N^2 \ .$$

$\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, \tau, c_{1,\tau}, \ldots, c_{n,\tau})$. The aggregator obtains the sum X_τ for time period τ by first computing $V_t := H(\tau)^{s_0} \prod_{i=1}^{n} c_{i,\tau} \bmod N^2$ and next X_τ as $X_\tau = \frac{V_\tau - 1}{N}$.

[Notice that since $(1 + N)$ has order N and $(1 + x_{i,\tau}N) \equiv (1 + N)^{x_{i,\tau}} \pmod{N^2}$, the so-obtained value for X_τ is defined modulo $M = N$.]

B. DEFERRED PROOF OF OUR DCR INSTANTIATION

The proof of our DCR instantiation in Section 5.3 (Theorem 5.1) is similar to the proof of Theorem 4.3 but not exactly the same, since the scheme is not key-uniform (as hashing keys have multiple representations), and hk_0 is the opposite of the sum of the hk_i over the integers and not in the group $\mathcal{K} = \mathbb{Z}_{N\phi(N)}$.

We again do a proof by games. We use the same first four games: Game 0 to Game 3. Then, as the SPHF is not key-uniform, we just skip Game 4 and go directly to Game 5 (except now, in this game, we suppose that $\mathsf{hk} \xleftarrow{R} \{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$ and that $\mathsf{hk}_0 = \sum_{i=1}^{n} \mathsf{hk}_i$ over the integers). The problem, is that, when $i^\star \neq 0$, it is no more true that Game 6 is indistinguishable from Game 5: $\mathcal{B}$ cannot pick hk_i independently and uniformly at random in $\mathcal{K}$ (or even in $\{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$), for $i \in \{0, \ldots, n\} \setminus \{i^\star\}$ and cannot compute $\mathsf{hk}_{i^\star}$ as $-\sum_{i \in \{0,\ldots,n\} \setminus \{i^\star\}} \mathsf{hk}_i$, because hk_0 is not uniform in $\mathcal{K}$ nor in $\{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$, as it is the sum of the hk_i over the integers.

We remark however, that if $i^\star = 0$, the adversary nevers sees hk_0, and if in Game 6, $\mathcal{B}$ picks hk_i independently and uniformly at random in $\{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$), for $i \in \{0, \ldots, n\} \setminus \{i^\star\}$ and computes $\mathsf{hk}_{i^\star} = \mathsf{hk}_0$ as $-\sum_{i \in \{0,\ldots,n\} \setminus \{i^\star\}} \mathsf{hk}_i$ over the integers, then the proof still works and $Adv_5 \leq Adv_6 + 2n\varepsilon_s \leq n/2^\kappa$.

We now focus on the difficult case $i^\star \neq 0$.

We use the following Game 6: $\mathcal{B}$ picks $\mathsf{hk}_1, \ldots, \mathsf{hk}_n$ uniformly at random in $\{-2^\kappa N^2, \ldots, 2^\kappa N^2\}$ and set $\mathsf{hk}_0 = -\sum_{i \in \{1,\ldots,n\}} \mathsf{hk}_i$ over the integers (as in Game 5). However, as in the original Game 6 and contrary to Game 5, $\mathcal{B}$ computes $h_{i^\star,\tau}$ as $\prod_{i \in \{0,\ldots,n\} \setminus \{i^\star\}} h_{i,\tau}^{-1}$ (which is perfectly indistinguishable). Furthermore, for any $i \in \mathcal{S}^\star \setminus \{i^\star\}$, $\mathcal{B}$ computes $h_{i,\tau^\star}$ as $\mathsf{Hash}(\mathsf{hk}'_i, y_{\tau^\star})$, with hk'_i a random hashing key in $\mathcal{K}$ corresponding to the same projection key hp_i as hk_i, or in other words:

$$\mathsf{hk}_i \bmod \phi(N) = \mathsf{hp}_i = \mathsf{hk}'_i \bmod \phi(N)\,,$$

but $\mathsf{hk}'_i \bmod N$ and $\mathsf{hk}_i \bmod N$ are independent (and uniformly random). We remark that $h_{i,\tau^\star} = \mathsf{Hash}(\mathsf{hk}'_i, y_{\tau^\star}) = y_{\tau^\star}^{\mathsf{hk}'_i}$ could also be computed as (in the case of the SPHF we consider): $H_{i,\tau^\star} = \mathsf{Hash}(\mathsf{hk}_i, y_{\tau^\star}) \cdot H'_i$ with $H'_i \xleftarrow{R} \Pi'$. Therefore, Game 6 (in this proof) is actually similar to Game 7 in the original proof, and $\mathcal{A}$ has no advantage in this game: $Adv_6 = 0$.

To conclude the proof, we just need to prove that Game 6 is statistically indistinguishable from Game 5. We suppose w.l.o.g. that $\mathcal{S}^\star \setminus \{i^\star\} = \{1, \ldots, m'\}$, where $m' = m < n$ if $i^\star \notin \mathcal{S}^\star$ and $m' = m - 1 < n$ otherwise. We also suppose that $i^\star = n$. We first remark that for any $i \in \mathcal{S}^\star \setminus \{i^\star\}$ and for any time period $\tau \neq \tau^\star$, $H_{i,\tau}$ can be computed from $\mathsf{hp}_i = \mathsf{hk}_i \bmod \phi(N)$ and hk_i is only used to compute hk_0. Define the following probability

distributions (for $j = 0, \ldots, m$):

$$\mathcal{D}_j = \Big\{ (\mathsf{hk}_0, \mathsf{hp}_1, \mathsf{hk}'_1, \ldots, \mathsf{hp}_{m'}, \mathsf{hk}'_{m'}, \mathsf{hk}_{m'+1}, \ldots, \mathsf{hk}_{n-1}) \mid \\ \mathsf{hk}_1, \ldots, \mathsf{hk}_n \xleftarrow{R} \{-2^\kappa N^2, \ldots, 2^\kappa N^2\}, \\ \mathsf{hk}_0 \leftarrow -\sum_{i=1}^{n} \mathsf{hk}_i, \mathsf{hk}'_1 \xleftarrow{R} \mathcal{K}, \ \ldots, \ \mathsf{hk}'_j \xleftarrow{R} \mathcal{K}, \\ \mathsf{hk}'_{j+1} = \mathsf{hk}_{j+1} \bmod N\phi(N), \ \ldots, \ \mathsf{hk}'_{m'} = \mathsf{hk}_{m'} \bmod N\phi(N) \text{ such that} \\ \mathsf{hp}_1 = \mathsf{hk}_1 \bmod \phi(N) = \mathsf{hk}'_1 \bmod \phi(N), \ \ldots, \\ \mathsf{hp}_{m'} = \mathsf{hk}_{m'} \bmod \phi(N) = \mathsf{hk}'_{m'} \bmod \phi(N) \Big\} \ .$$

When all these variables are drawn from distibution $\mathcal{D}_0$, we get Game 5, while when they are drawn from distribution $\mathcal{D}_m$, we get Game 6.

To do that, we use the following lemma:

LEMMA B.1. *Let M, N', N'' be three integers, with $M \geq N'N''$ and N' coprime with N''. Let X, Y be two random uniform random variables in $\{-M, \ldots, M\}$, and X' be a uniform random variable in $\{0, \ldots, N'-1\}$. If we suppose that X, Y, X' are mutually independent, then the statistical distance between the distribution of $(X \bmod N', X \bmod N'', X+Y)$ and of $(X', X \bmod N'', X+Y)$ is at most $\frac{N''(N'+1)}{2(2M+1)}$.*

PROOF OF LEMMA B.1. This statistical distance is

$$\frac{1}{2} \sum_{\substack{x' \in \{0,\ldots,N'-1\} \\ x'' \in \{0,\ldots,N''-1\} \\ z \in \{-2M,\ldots,2M\}}} \Big| \Pr[X \bmod N' = x', X \bmod N'' = x'', X+Y = z] - \\ \Pr[X' = x', X \bmod N'' = x'', X+Y = z] \Big| \ .$$

We have X', X, Y are mutually independent, so

$$\begin{aligned} &\Pr[X' = x', X \bmod N'' = x'', X+Y = z] \\ &\quad = \sum_{\substack{x \in \{-M,\ldots,M\} \\ \text{such that } x \bmod N'' = x''}} \Pr[X' = x'] \cdot \Pr[X \bmod N'' = x, X+Y = z] \\ &\quad = \sum_{\substack{x \in \{-M,\ldots,M\} \\ \text{such that } x \bmod N' = x'}} \frac{1}{N'} \cdot \Pr[X \bmod N'' = x'', Y = z - x] \\ &\quad = \frac{n_{x'}}{N'(2M+1)^2} \end{aligned}$$

where $n_{x'}$ is the number of values $x \in \{-M, \ldots, M\}$ such that $x \bmod N'' = x''$ and $z - x \in \{-M, \ldots, M\}$. There are $2M + 1 - |z|$ values x such that $z - x \in \{-M, \ldots, M\}$, namely $-M, \ldots, M + z$ if $z \leq 0$, and $-M + z, \ldots, M$ if $z \geq 0$. Among these values either $\lfloor (2M + 1 - |z|)/N'' \rfloor$ or $\lfloor (2M + 1 - |z|)/N'' \rfloor + 1$ of them are such that $x \bmod N' = x'$. Therefore, we have:

$$\frac{(2M+1-|z|)}{N''} - 1 \leq n_{x'} \leq \frac{(2M+1-|z|)}{N''} + 1.$$

Moreover, we have

$$\begin{aligned}
&\Pr[X \bmod N' = x',\, X \bmod N'' = x'',\, X + Y = z] \\
&\quad = \sum_{\substack{x \in \{-M,\ldots,M\} \\ \text{such that } x \bmod N' = x' \\ \text{and } x \bmod N'' = x''}} \Pr[X \bmod N' = x',\, X \bmod N'' = x'',\, X + Y = z] \\
&\quad = \sum_{\substack{x \in \{-M,\ldots,M\} \\ \text{such that } x \bmod N' = x' \\ \text{and } x \bmod N'' = x''}} \Pr[X \bmod N' = x',\, X \bmod N'' = x'',\, Y = z - x] \\
&\quad = \frac{n_{x',x''}}{(2M+1)^2}
\end{aligned}$$

where $n_{x',x''}$ is the number of values $x \in \{-M,\ldots,M\}$ such that $x \bmod N' = x'$ and $x \bmod N'' = x''$ and $z - x \in \{-M,\ldots,M\}$. There are $2M + 1 - |z|$ values x such that $z - x \in \{-M,\ldots,M\}$, namely $-M,\ldots,M+z$ if $z \leq 0$, and $-M+z,\ldots,M$ if $z \geq 0$. Among these values either $\lfloor (2M+1-|z|)/(N'N'') \rfloor$ or $\lfloor (2M+1-|z|)/(N'N'') \rfloor + 1$ of them are such that $x \bmod N' = x'$ and $x \bmod N'' = x''$, as N' and N'' are coprime, and thanks to the CRT. Therefore, we have:

$$\frac{(2M+1-|z|)}{N'N''} - 1 \leq n_{x',x''} \leq \frac{(2M+1-|z|)}{N'N''} + 1 \ .$$

Thus, the statistical distance satisfies

$$\begin{aligned}
\frac{1}{2} \sum_{x',x'',z} \left| \frac{n_{x'}}{N'(2M+1)^2} - \frac{n_{x',x''}}{(2M+1)^2} \right| &= \frac{1}{2} \sum_{x',x'',z} \frac{1}{(2M+1)^2} \left| \frac{n_{x'}}{N'} - n_{x''} \right| \\
&\leq \frac{1}{2} \sum_{x',x'',z} \frac{1}{(2M+1)^2} \left(1 + \frac{1}{N'} \right) \\
&\leq \frac{N''(N'+1)}{2(2M+1)} \ .
\end{aligned}$$

This concludes the proof of the lemma. □

We use the lemma with $M = 2^\kappa N^2$, $N' = N$, and $N'' = \phi(N)$. For any integers $x' \in \{0,\ldots,N'-1\}$ and $x'' \in \{0,\ldots,N''-1\}$, let us denote by $x = \mathsf{CRT}(x', x'')$ the unique integer $x \in \{0,\ldots,N'N''-1\}$ such that $x \bmod N' = x'$ and $x \bmod N'' = x''$. Then, in $\mathcal{D}_{j-1}$, we remark that hk_j, hp_j, hk'_j, hk_n, and hk_0 can alternatively be defined as:

$$\begin{aligned}
&\mathsf{hk}_j = X \xleftarrow{R} \{-M,\ldots,M\} \text{ (not actually used in the distribution directly)} \\
&\mathsf{hp}_j = X \bmod N'' \\
&\mathsf{hk}'_j = \mathsf{hk}_j \bmod N'N'' = \mathsf{CRT}(X \bmod N',\, X \bmod N'') \\
&\mathsf{hk}_n = Y \xleftarrow{R} \{-M,\ldots,M\} \\
&\mathsf{hk}_0 = -\sum_{i \in \{1,\ldots,n-1\} \setminus \{j\}} \mathsf{hk}_i - (X+Y)\,,
\end{aligned}$$

while in $\mathcal{D}_j$, these random variables can alternatively be defined as:

$$\begin{aligned}
\mathsf{hk}_j &= X \xleftarrow{R} \{-M, \ldots, M\} \text{ (not actually used in the distribution directly)} \\
\mathsf{hp}_j &= X \bmod N'' \\
\mathsf{hk}'_j &= \mathsf{hk}_j \bmod N'N'' = \mathsf{CRT}(X',\, X \bmod N'') \text{ where } X' \xleftarrow{R} \{0, \ldots, N'-1\} \\
\mathsf{hk}_n &= Y \xleftarrow{R} \{-M, \ldots, M\} \\
\mathsf{hk}_0 &= - \sum_{i \in \{1,\ldots,n-1\} \setminus \{j\}} \mathsf{hk}_i - (X+Y) \ .
\end{aligned}$$

Back to the theorem, thanks to Lemma B.1, we can conclude $\mathcal{D}_{j-1}$ and $\mathcal{D}_j$ are $(N'+1)N''/(2(2M+1))$-close, and the statistical distance between $\mathcal{D}_0$ and $\mathcal{D}_n$ is at most:

$$n \frac{(N'+1)N''}{2(2M+1)} \leq \frac{n}{2^{\kappa+1}} \ .$$

Therefore, $Adv_5 \leq Adv_6 + n/2^\kappa$. □

C. IMPOSSIBILITY RESULT OF TIGHTNESS FOR A PREVIOUS SCHEME

We now show that any blackbox non-rewinding reduction from the aggregator obliviousness of Shi *et al.*'s scheme in [2011] to a non-interactive problem loses a factor of at least n^2. This bound cannot be improved in the Shi *et al.*'s scheme. This impossibility result is shown by outlining a meta-reduction (that is, a reduction against the reduction) as in [Coron 2002]. The idea is to show that any reduction losing a factor better than (about) n^2 can be converted into an adversary for the original hard problem, by constructing an adversary $\mathcal{B}$ which acts as an adversary for the reduction but which can also rewind the reduction.

The basic idea is that, in the scheme of Shi *et al.*, sk_i is completely defined (from the information theory viewpoint) when a ciphertext $c_{i,1}$ for 0 (for example) is given; and sk_0 is defined by $(c_{i,1})_{i \in \{1,\ldots,n\}}$.

More precisely, suppose that the reduction can solve the original hard problem with probability ε_R, when playing with any adversary breaking the aggregator obliviousness with advantage $2\varepsilon_{\mathcal{A}} - 1$. We construct an adversary $\mathcal{B}$ (which has the right to rewind the reduction) as follows: $\mathcal{B}$ will first ask for the aggregator secret key sk_0 and for ciphertexts $c_{i,1}$ of 0 for time period 1 (and for each user i). It can check that any secret key sk_i given by the reduction is valid or not (for each user i), with respect to $c_{i,1}$ (by checking that $\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, 1, 0) = c_{i,\tau}$) and that sk_0 is valid (by checking that $\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, 1, c_{1,1}, \ldots, c_{n,1}) = 0$). Then it will choose two random users $i_1 \neq i_2$. For all pairs of distinct users $\{i_3, i_4\} \neq \{i_1, i_2\}$, it then asks all the secret keys sk_i (in an arbitrary order) except i_3 and i_4, store them, and then rewind the adversary just after the corruption of sk_0 —$\mathcal{B}$ therefore rewinds the reduction $(n-1)(n-2)/2$ times (one for each pair $\{i_3, i_4\} \neq \{i_1, i_2\}$). After that, $\mathcal{B}$ will have stored $(n-1)(n-2)/2$ keys sk_{i_1} and $(n-1)(n-2)/2$ keys sk_{i_2}. If any of them are invalid, $\mathcal{B}$ aborts and returns $b' = 0$ with probability $1/2$, and $b' = 1$ otherwise.

Otherwise, $\mathcal{B}$ then asks all the secret keys sk_i except for i_1 and i_2. If none of them are valid, $\mathcal{B}$ aborts and returns $b' = 0$ with probability $1/2$, and $b' = 1$ otherwise. Otherwise, it just submits the challenge: $\mathcal{S}^\star = \{i_1, i_2\}$, $\tau^\star = 2$, $(x^{(0)}_{i_1,2} = 0, x^{(0)}_{i_2,2} = 1)$ and $(x^{(1)}_{i_1,2} = 1, x^{(1)}_{i_2,2} = 0)$. The reduction will returns a pair of ciphertexts $\langle c_{i_1,1}, c_{i_1,2} \rangle$ encrypting either $\langle x^{(0)}_{i_1,2}, x^{(0)}_{i_2,2} \rangle$ or $\langle x^{(1)}_{i_1,2}, x^{(1)}_{i_2,2} \rangle$. And $\mathcal{B}$ will check that these ciphertexts are coherent with sk_0, by computing $c_{i,2} = \mathsf{Enc}(\mathsf{param}, \mathsf{sk}_i, 2, 0)$ for all $i \notin \{i_1, i_2\}$, and checking that $\mathsf{AggrDec}(\mathsf{param}, \mathsf{sk}_0, 2, c_{1,2}, \ldots, c_{n,2}) = 1$. Finally, if $\mathcal{B}$ got a valid secret key sk_{i_1} and if $\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_{i_1}, 2, 0) = c_{i_1,2}$, $\mathcal{B}$ sets $b'' = 0$; if $\mathcal{B}$ got a valid secret key sk_{i_2} and

$\mathsf{Enc}(\mathsf{param}, \mathsf{sk}_{i_2}, 2, 1) = c_{i_2,1}$, $\mathcal{B}$ sets $b'' = 0$; otherwise, $\mathcal{B}$ sets $b'' = 1$. And $\mathcal{B}$ outputs $b' = b''$ with probability $\varepsilon_{\mathcal{A}}$ and $b' = 1 - b''$ otherwise.

We remark that if sk_{i_1} or sk_{i_2} is valid, $\mathcal{B}$ would behave exactly as an all powerful (non polynomially bounded) adversary $\mathcal{A}$ which would do the same as $\mathcal{B}$, except it does not perform any rewinding (and so never corrupts sk_{i_1} and sk_{i_2}), and instead computes sk_{i_1} and sk_{i_2} simply by trying all possible values. This is true thanks to the check with sk_0 which ensures that if $c_{i_1,2}$ is an encryption of 0 (respectively 1), then $c_{i_2,2}$ is an encryption of 1 (respectively 0), and so knowing only one of sk_{i_1} and sk_{i_2} is sufficient. In addition, if any sk_i for some $i \notin \{i_1, i_2\}$ (obtained after the last rewinding for $\mathcal{B}$) is not valid, both $\mathcal{A}$ and $\mathcal{B}$ abort. Therefore, $\mathcal{A}$ and $\mathcal{B}$ behave identically, except if all secret keys sk_i (for $i \notin \{i_1, i_2\}$) are valid (after the last rewinding for $\mathcal{B}$) but $\mathcal{B}$ cannot find a valid key sk_{i_1} or a valid key sk_{i_2} among all keys it got from all the rewindings. We call 'bad case' the case where the previous bad event happens, and let ε_{bad} the probability of this event.

The advantage of $\mathcal{A}$ (to break the aggregator obliviousness) is exactly $Adv_{\mathcal{A}} = 2\varepsilon_{\mathcal{A}} - 1$, since when $\mathcal{A}$ plays against the real challenger for the aggregator obliviousness, $\mathcal{A}$ never aborts, and the bit b'' computed by $\mathcal{A}$ is always equal to b, the bit chosen in the game, hence $b' = b$ with probability $\varepsilon_{\mathcal{A}}$. In the bad case, $\mathcal{B}$ outputs $b' = b$ with probability $1/2$ instead of $\varepsilon_{\mathcal{A}}$ for $\mathcal{A}$, while outside the bad case, $\mathcal{B}$ and $\mathcal{A}$ output $b' = b$ with the same probability. Therefore, when playing with $\mathcal{B}$, the reduction solves the original hard problem with probability at least:

$$\varepsilon_R - \varepsilon_{\text{bad}} \cdot |\varepsilon_{\mathcal{A}} - 1/2| = \varepsilon_R - \varepsilon_{\text{bad}} \cdot Adv_{\mathcal{A}}/2$$

which in term of advantage (if the reduction solves a decisional problem) is

$$2\left(\varepsilon_R - \varepsilon_{\text{bad}} \left|\varepsilon_{\mathcal{A}} - \tfrac{1}{2}\right|\right) - 1 = (2\varepsilon_R - 1) - \varepsilon_{\text{bad}}|2\varepsilon_{\mathcal{A}} - 1| = Adv_R - \varepsilon_{\text{bad}} \cdot Adv_{\mathcal{A}} \ .$$

This has to be negligible, otherwise the hard problem would not be hard. Therefore, ε_R or Adv_R cannot be larger than $\varepsilon_{\text{bad}} \cdot Adv_{\mathcal{A}}/2$ or $\varepsilon_{\text{bad}} \cdot Adv_{\mathcal{A}}$ (minus some negligible factor in the security parameter). We just need to prove that $\varepsilon_{\text{bad}} \geq 2/(n(n-1))$, to show our impossibility result.

For that purpose, for any integers j_1 and j_2, let E_{j_1,j_2} denote the event that, when the secret keys sk_i for $i \in \{1, \ldots, n\} \setminus \{j_1, j_2\}$ are corrupted, the secret keys given by the reduction are all valid. We clearly have:

$$\varepsilon_{\text{bad}} \leq \Pr\left[E_{i_1,i_2} \wedge \left(\bigwedge_{\{i_3,i_4\} \neq \{i_1,i_2\}} \neg E_{i_3,i_4}\right)\right],$$

since if E_{i_3,i_4} is true for some $\{i_3, i_4\} \neq \{i_1, i_2\}$ ($i_3 \neq i_4$), then $i_1 \notin \{i_3, i_4\}$ or $i_2 \notin \{i_3, i_4\}$, and we get a valid secret key sk_{i_1} or sk_{i_2}. If we fix everything except the choice of i_1 and i_2, each event E_{i_3,i_4} is either satisfied (it has probability 1) or not satisfied (it has probability 0). If two distinct event E_{i_5,i_6} and $E_{i'_5,i'_6}$ are satisfied, the event $\bigwedge_{\{i_3,i_4\} \neq \{i_1,i_2\}} \neg E_{i_3,i_4}$ never happens and $\varepsilon_{\text{bad}} = 0$. If no event is satisfied, the event E_{i_1,i_2} never happens and $\varepsilon_{\text{bad}} = 0$. Finally, if only one event E_{i_5,i_6} is satisfied, $E_{i_1,i_2} \wedge (\bigwedge_{\{i_3,i_4\} \neq \{i_1,i_2\}} \neg E_{i_3,i_4})$ happens only when $\{i_1, i_2\} = \{i_5, i_6\}$, which happens with probability $2/(n(n-1))$. Therefore, $\varepsilon_{\text{bad}} \leq 2/(n(n-1))$.