Noname manuscript No.
(will be inserted by the editor)

Secure and Collusion Resistant Data Aggregation from Convertible Tags

Iraklis Leontiadis · Ming Li

Received: date / Accepted: date

Abstract The progress in communication and hardware technology increases the computational capabilities of personal devices. Aggregators, acting as third parties, are interested in learning a statistical function as the sum over a census of data. Users are reluctant to reveal their information in cleartext, since it is treated as personal sensitive information. The paradoxical paradigm of preserving the privacy of individual data while granting an untrusted third party to learn in cleartext a function thereof, is partially addressed by the current privacy preserving aggregation protocols.

Current solutions are either focused on an honest-but-curious Aggregator who is trusted to follow the rules of the protocol or they model a malicious Aggregator with trustworthy users. In this paper we are the first to propose a protocol with fully malicious users who collude with a malicious Aggregator in order to forge a message of a trusted user. We introduce the new cryptographic primitive of *convertible tag*, that consists of a two-layer authentication tag. Users first tag their data with their secret key and then an untrusted *Converter* converts the first layer tags in a second layer. The final tags allow the Aggregator to produce a proof for the correctness of a computation over users' data. Security and privacy of the scheme is preserved against the *Converter* and the Aggregator, under the notions of *Aggregator obliviousness* and *Aggregate unforgeability* security definitions, augmented with malicious users. Our protocol is provably secure and experimental evaluations demonstrate its practicality.

Keywords data privacy · data security · convertible tags · collusion resistant aggregation

Electrical and Computer Engineering Department
University of Arizona, USA
E-mail: leontiad,lim@email.arizona.edu

1 Introduction

The folklore model of Alice and Bob who want to exchange messages in a secure way, has been extensively analyzed. Nowadays, with the progress of communication and computing technology, users are able to produce big amount of data, which is shared with untrusted parties. As such, the idea of locally holding the data is of the past. Users leverage the computational and storage capabilities in order to store and analyze their data. Solutions tailored for this scenario propose a new model for outsourced data computations. In the paper we are focused on secure aggregation. In a nutshell, in an aggregation protocol, untrusted parties collect individual users' data in order to compute a function over their cleartext data. The paradigm of data collection and analysis is motivated by a plethora of real world scenarios:

- Smart metering data is collected by an energy supplier in order to perform energy forecasting for cost minimizations. On the other hand users want to protect their individual privacy and apply a privacy preserving mechanism on their data.
- In a healthcare scenario patients leave their health traces to hospitals. These traces comprise health care sensitive data and a compromise thereof, affect negatively the patients: A hospital which acts as a data enclave for patients data may collude with an insurance company. The latter may decline an insurance subscription to a patience according its health care data.

In the aforementioned use cases an untrusted Aggregator computes in cleartext a function f over users' data and forwards the result to a Data Analyzer. The paradox stems from the desire of each user to protect

its individual privacy while the Aggregator wants to learn in cleartext f over users' data. Existing literature is focused either on protecting individual privacy [16, 34, 37, 45] or on improving the security model with a malicious Aggregator; who will try to convince a Data Analyzer, who acts as an honest verifier, that the result of computations comes from genuine data inputs. In [38] the authors by leveraging the encryption scheme of Shi *et al.* [45] they enrich the typical data collection and analysis protocol with a proof computed by a malicious Aggregator, which allows the Data Analyzer to verify the correctness of computations. However the authors employ a rather weak model. During their analysis, users are assumed as trustworthy and they do not collude with the Aggregator. However, this assumption is not realistic in a real world scenario in which trustworthiness is not guaranteed. Namely, users can collude with the Aggregator in order to change the protocol's messages at their need. This would have devastating consequences on users' security. In [26] the authors propose a solution in which trustworthiness of users is correlated with the validity of their produced data. Their solutions incorporates a blind commitment before the collection of the data. In between the commitment and the aggregation phase users cannot change their data. However a malicious user is able to alter its real data before the commitment phase, thus violating the validity of data.

We propose a secure aggregation protocol in the presence of untrustworthy users. In this setting users are allowed to collude with a malicious Aggregator, without affecting the unforgeability of the scheme. We only require that users send correct data and not fake information. The striking attribute of our protocol which is of independent importance is a new cryptographic primitive named *convertible tag*. Users tag their data with a convertible tag using independent randomness. This allows users to collude with a malicious Aggregator without the latter being able to forge user's data. The tag is convertible, in the sense that a semi-trusted third party with some auxiliary information computed by each user, can convert it to a another tag, which is able to be aggregated with respect to the function $f = \sum$. Informally, the security guarantees for *convertible tags* assure that any collusion of the user with a malicious Aggregator cannot forge non-genuine data, originating from other users. Plugging convertible tags to a secure aggregation protocol also assures unforgeability of data aggregation as formalized in [38] and Aggregator obliviousness [45]. That is, a malicious Aggregator cannot convince an honest verifier for the correctness of computation f that arises from non-genuine data inputs and moreover individual privacy of users' inputs is preserved thanks to the obliviousness property. We summarize the contributions of this paper as follows.

Contributions

- In the aim of assuring collusion resistant aggregation we come up with the cryptographic primitive of *convertible tag*. Users can choose independently their tag keys. The tags are unified under common randomness with the aid of a semi-honest third party, called hereafter the *Converter*. The convertible tags assure *obliviousness* of computations against a malicious Aggregator and a semi-honest *Converter*, without jeopardizing unforgeability.
- We extend the current security definitions of secure aggregation protocols with collusions: a) between users and Aggregator, b) between users and the *Converter*, c) between the Aggregator and the *Converter*. Our protocol is provably secure under standard assumptions in the random oracle model.
- Thanks to our construction, the protocol achieves constant time symmetric verification in a multi-user setting.

Outline In section 2 we introduce the problem this paper addresses and we identify the lack of a stronger security definition from existing protocols. Afterwards, in section 3 we review similar cryptographic primitives with *convertible tags*. We continue in section 4 with the core idea of our solution. Before we present our solution, we sketch the preliminaries in section 5. The protocol is presented in full details in section 6. Section 7 demonstrates the security of our solution and in section 8 we analyze the overhead of our scheme. Section 9 compares our solution with state of the art schemes. Finally, we conclude in section 10.

2 Problem Statement

For a secure aggregation protocol, we assume a set of n users $\mathbb{U} = \{\mathcal{U}_i\}_{i=1}^{n}$, each one producing time series personal data inputs $x_{i,t}$. Users encrypt their data with an encryption algorithm, which produces ciphertexts $c_{i,t}$. Ciphertexts are collected by an Aggregator $\mathcal{A}$, whose main goal is to learn a function $f = \sum x_{i,t}$ in cleartext over users' data and forward the result to a trustworthy Data Analyzer $\mathcal{DA}$, who does not communicate with each user. We assume a malicious Aggregator who does not follow the rules of the protocol and seeks to infer more information from the exchanged messages of the protocol. More specifically the Aggregator will try to convince an honest verifier $\mathcal{DA}$ for the correctness of computations over non-genuine data. To protect against the malicious Aggregator users further tag their

data in such a way that a proof of correct computations can be constructed by the Aggregator and will convince the verifier.

We recall in this section the syntax of a secure aggregate protocol as described in [38].

2.1 Syntax

- $\mathsf{Setup}(1^\lambda) \rightarrow (\mathsf{pp}, \mathsf{sk_A}, \{\mathsf{sk}_i\}_{\mathcal{U}_i,\in\mathbb{U}}, \mathsf{vk})$: It is a randomized algorithm run by a trusted dealer $\mathcal{KD}$, which on input of a security parameter λ outputs the public parameters pp that will be used by subsequent algorithms, the Aggregator $\mathcal{A}$'s secret key $\mathsf{sk_A}$, the secret keys $\mathsf{sk_i}$ of users $\mathcal{U}_i$ and the public verification key vk.
- $\mathsf{EncTag}(t, \mathsf{sk}_i, x_{i,t}) \rightarrow (c_{i,t}, \mathsf{st}_{i,t})$: It is a randomized algorithm which on inputs of time interval t, secret key sk_i of user $\mathcal{U}_i$ and data $x_{i,t}$, encrypts $x_{i,t}$ to get a ciphertext $c_{i,t}$ and computes a tag $\mathsf{st}_{i,t}$ that authenticates $x_{i,t}$.
- $\mathsf{Aggregate}(\mathsf{sk_A}, \{c_{i,t}\}_{\mathcal{U}_i\in\mathbb{U}}, \{\mathsf{st}_{i,t}\}_{\mathcal{U}_i\in\mathbb{U}}) \rightarrow (\mathsf{sum_t}, \sigma_\mathsf{t})$: It is a deterministic algorithm run by the Aggregator $\mathcal{A}$. It takes as inputs Aggregator $\mathcal{A}$'s secret key $\mathsf{sk_A}$, ciphertexts $\{c_{i,t}\}_{\mathcal{U}_i\in\mathbb{U}}$ and authentication tags $\{\mathsf{st}_{i,t}\}_{\mathcal{U}_i\in\mathbb{U}}$, and outputs in cleartext the sum $\mathsf{sum_t}$ of the values $\{x_{i,t}\}_{\mathcal{U}_i\in\mathbb{U}}$. Moreover, it computes a proof σ_t assessing the correctness of $\mathsf{sum_t}$, using the authentication tags $\{\mathsf{st}_{i,t}\}_{\mathcal{U}_i\in\mathbb{U}}$.
- $\mathsf{Verify}(\mathsf{vk}, \mathsf{sum_t}, \sigma_\mathsf{t}) \rightarrow \{0, 1\}$: It is a deterministic algorithm that is executed by the Data Analyzer $\mathcal{DA}$. It outputs 1 if Data Analyzer $\mathcal{DA}$ is convinced that the sum $\mathsf{sum_t} = \sum_{\mathcal{U}_i\in\mathbb{U}}\{x_{i,t}\}$; and 0 otherwise, with the aid of the verification key vk.

2.2 Security Model

We build upon the model as presented in [38] and we further assume that users are not trustworthy. We only require that each user submits real data and not fake inputs. Notably, users can collude with the Aggregator in order to forge non-genuine tags for a legitimate user. This has a negative result on the scheme's security, since the security definition of *aggregate unforgeability* is not assured anymore. In a nutshell, *aggregate unforgeability* definition follows the classical message tag unforgeability under chosen message attack, with the difference that adversary $\mathcal{A}$ cannot forge an aggregate tag with respect to the computation f. That is, if users submit tags $\mathsf{st}_{i,t}$ for their private data inputs $x_{i,t}$ then $\mathcal{A}$ can only compute a valid aggregate tag st_t for the sum computation over $x_{i,t}$ and nothing else. We show how the scheme in [38] does not assure *aggregate unforgeability* in the presence of non-legitimate users, who collude with a malicious Aggregator $\mathcal{A}$. A malicious user $\mathcal{U}_m$ shares the secret information a with the Aggregator during the protocol execution for an arbitrary time interval t. The Aggregator $\mathcal{A}$ then can forge another user's tag with a as follows: After obtaining $\mathsf{st}_{l,t} = H(t)^{\mathsf{tk}_l}(g_1^a)^{x_{l,t}}$ from a legitimate user $\mathcal{U}_l$ at time interval t, $\mathcal{A}$ computes $\mathsf{st}_{l,t} \cdot (g_1^a)^v = H(t)^{\mathsf{tk}_l}(g_1^a)^{x_{l,t}+v}$, for a value v of its choice. Afterwards, when the Data Analyzer asks for the sum computation at the time interval the malicious Aggregator can claim for a non-genuine sum$= \sum_{i=1}^n x_{i,t} + v$ instead of the correct sum $\sum_{i=1}^n x_{i,t}$ along with a sound proof, which would be verified by the Data Analyzer: During the verification phase the Data Analyser verifies:

$$e(\sigma_\mathsf{t}, g_2) \stackrel{?}{=} e(H(t), \mathsf{vk}_1)e(g_1^{\mathsf{sum_t}}, \mathsf{vk}_2)$$

, where $\sigma_\mathsf{t} = \prod_{i=1}^n \mathsf{st}_{i,t}$. $\mathcal{A}$ can compute σ_t' as follows: $\sigma_\mathsf{t}' = \prod_{i=1\wedge i\neq l}^n \mathsf{st}_{i,t} \cdot H(t)^{\mathsf{tk}_l}(g_1^a)^{x_{l,t}+v}$. Finally it sends to the Data Analyzer $\mathsf{sum_t}' = \sum_{i=1}^n x_{i,t} + v, \sigma_\mathsf{t}'$. The latter correctly verifies as:

$$\begin{aligned}
e(\sigma_\mathsf{t}', g_2) &= e(\prod_{i=1\wedge i\neq l}^{n} \mathsf{st}_{i,t} \cdot H(t)^{\mathsf{tk}_l}(g_1^a)^{x_{l,t}+v}, g_2) \\
&= e(H(t)^{\mathsf{tk}_l}(g_1^a)^{x_{l,t}+v} \cdot \prod_{i=1\wedge i\neq l}^{n} H(t)^{\mathsf{tk}_i} g_1^{a x_{i,t}}, g_2) \\
&= e(H(t)^{\sum_{i=1}^n \mathsf{tk}_i} g_1^{a\sum_{i=1}^n x_{i,t}+v}, g_2) \\
&= e(H(t)^{\sum_{i=1}^n \mathsf{tk}_i}, g_2)e(g_1^{a\sum_{i=1}^n x_{i,t}+v}, g_2) \\
&= e(H(t), g_2^{\sum_{i=1}^n \mathsf{tk}_i})e(g_1^{\sum_{i=1}^n x_{i,t}+v}, g_2^a) \\
&= e(H(t), g_2^{\sum_{i=1}^n \mathsf{tk}_i})e(g_1^{\mathsf{sum_t}'}, g_2^a) \\
&= e(H(t), \mathsf{vk}_1)e(g_1^{\mathsf{sum_t}'}, \mathsf{vk}_2)
\end{aligned}$$

. Thus, $\mathcal{A}$ can produce a valid proof by aggregating all tags and the forged one, for a sum that comes from non-genuine data. We also inherit the privacy definitions of Aggregator obliviousness, which protects individual privacy. A malicious Aggregator from the computation of the sum in cleartext over individual data inputs cannot jeopardize individual privacy. The privacy definition is expressed as a security game.

3 Related work

Similar cryptographic primitives have been proposed in the literature for the purpose of unforgeability with privacy. Blind signatures provide privacy by allowing the signer to sign a message blindly, without learning what it signs [18]. Group signatures [19] provide anonymity

by allowing any member of an authorized group to sign on behalf of the group manager. Group signatures provide traceability and non-frameability. The traceability property requires that no adversary can compute a signature that cannot be traced to a user and non-frameability assures that a malicious group manager cannot falsely accuse a user. With proxy signatures [9, 41] and its variations (anonymous [28], private [24, 32]), signing rights are delegated to a proxy who signs on behalf of a user. Proxy Re-Signatures (PRS) [3, 8, 40] translate signature for one party to another one. PRS share some properties with convertible tags. We carefully compare our new primitive with the aforementioned constructions below.

Blind signatures. Chaum first introduced the notion of *blind signatures* [17]. A user sends a blinded version of its message to the signer and the latter signs without learning the underlying message. The user then obtains the signature on the original message and sends the signature to the verifier. Apart from confidentiality, blind signatures guarantee [44] also anonymity and they are useful for a broad range of applications, as e-cash [11] and anonymous credentials [12]. Similarly with the *convertible tags* blind signatures offer privacy on top of authentication but only for the third party who signs and not for the verifier. The verifier in a blind signature verifies the correctness of a message in cleartext. In contrast, *convertible tags* extend this functionality with privacy, since there is not one-to-one message signature verification but verification of the correctness of an aggregate result over data. Moreover, in a multi-user setting, *convertible tags*, offer increased security, other than unlinkability, in case of collusions between a user and the signer. The user can verify the well-formness of the tag. In contrast blind signatures assume the signer as trusted to sign correctly.

Group signatures. Group signatures [2, 7, 10, 13–15] allow a member of a group to sign on behalf of a group manager in such a way that anonymity of the sender is preserved. Moreover they guarantee traceability of the signatures, non-frameability and coalition resistance. The model of *convertible tags* differ from group signatures in the sense that groups signatures do not offer confidentiality over the entire group messages and moreover they do not support homomorphic operations on the signatures.

Proxy signatures. In proxy signatures [9, 41] the signer delegates its signing rights to an authorized proxy. The proxy can sign on behalf of the designator and the receiver of the signatures can verify the authenticity of the signature as originated from the designator. In practice the secret key of the original signer is split between the receiver and the proxy. Variations of proxy signatures as warrant-signatures [24, 32] restrict the proxy to sign only specific parts of the messages without being able to learn the space of the allowed messages that it can sign. *Convertible* tags enable a multi-user setting, in which multiple tags from different users are converted in a single tag with common randomness.

Proxy re-signatures. The primitive of proxy re-signatures [3, 8, 40] allows a designator to delegate a transformation operation on its signature with the aid of proxy in order the latter to transform the original signature signed with the signature key of a different user. The proxy re-signature primitive bears similarities with the convertible tags primitive since in both there is a transformation mechanism by a third party, who converts the authentication tags. However convertible tags operate in a different model: multiple users tag their data such that the third party cannot learn the authenticated data. As such, confidentiality is being preserved in contrast with proxy-re signatures in which there is only authenticity guarantee. Another major issue with proxy re-signatures is that they are not homomorphic, while convertible are constructed not for a per message verification but for computation verification.

Conceptually, *convertible tags* can be viewed as a combination of blind signatures, group signatures, and proxy (re-) signatures. They employ the privacy guarantee of confidentiality of blind signatures, the communication model of groups signatures and the transformation property of a signature from one user to another as with proxy (re-) signatures. However a simple assembly of the aforementioned primitives for the construction of a *convertible tag* is not a trivial plug in of all those primitives, simply because in case of collusions the security guarantees of each are not preserved. The model of aggregation resembles e-voting schemes, in which the functionality f is the counting of votes. However our model differs from e-voting schemes [22] since in the latter privacy is preserved thanks to the secret sharing of the decryption key, while our technique allows the untrusted Aggegator to posses a secret decryption key in order to learn in cleartext the sum over a data sample, without compromising individual privacy. Notice also that unforgeable signatures on the tag solves the problem, but that would incur extra computational complexity to the Aggregator for verifying each signature, and the different public keys for all users burden its storage complexity.

4 Idea and Model

4.1 Idea

The core idea of our solution for collusion resistant aggregation is a symmetric authentication mechanism at the target group of bilinear pairings. Each user chooses uniformly at random tag keys for the authentication tag, which at a first level, is named metatag. Users send their metatags to a semi-honest party, the *Converter* $\mathcal{C}$ and their ciphertexts to the malicious Aggregator $\mathcal{A}$. Along with the metatags each user transmits to $\mathcal{C}$ some auxiliary information coupled with a blinded version of their secret tag key. $\mathcal{C}$ then couples all this information and ends up with the final tag of each user at the second level. The coupling annihilates the randomness per user and transforms the metatags to the final *convertible tag*, that is forwarded to the malicious Aggregator $\mathcal{A}$. Users upon receiving their tags from $\mathcal{C}$ validate its correctness. This is happening in order to ensure that in case of a collusion between a colluding user and the *Converter* $\mathcal{C}$, the latter cannot forward a forged tag, with the key that is used by $\mathcal{C}$ to couple the metatag and the auxiliary information. That is, a malicious user cannot extract the randomness used for the final computation of the authentication tag in case of collusion with the malicious $\mathcal{A}$, in order to forge an authentication tag for another user. Aggregator receives all tags and ciphertexts. $\mathcal{A}$ then decrypts and learns the result $\mathsf{sum_t} = f = \sum_{i=1}^{n} x_{i,t}$ and computes a proof of correct computations σ_{t} based on the convertible tags.

Finally, $\mathcal{A}$ forwards to the data analyzer $\mathcal{DA}$ the result $\mathsf{sum_t}$ and the proof σ_{t}. $\mathcal{DA}$ verifies the correctness of computations as an honest verifier in constant time. The *convertible tags* assure *Aggregator obliviousness* and *aggregate unforgeability*. In a nutshell with *Aggregator obliviousness* $\mathcal{A}$ cannot learn anything more than the aggregate result $\sum_{i=1}^{n} x_{i,t}$. *Aggregate unforgeability* guarantees the correct computation of $\mathsf{sum_t} = f = \sum_{i=1}^{n} x_{i,t}$. Both security guarantees are enriched, in contrast with previous work [38], with collusions between malicious users, Aggregator $\mathcal{A}$ and *Converter* $\mathcal{C}$. Thus, our solution assures :

1. Collusion resistance between a malicious user and a malicious Aggregator $\mathcal{A}$, thanks to the individual keys chosen by each user. Despite the convertible tags that in the end cancel out all the individual keys and use common randomness in order $\mathcal{A}$ to compute a proof of correctness based on the sum computation, individual randomness chosen by each user permits collusions between a user and an Aggregator without the latter being able to forge a tag of a legitimate user.
2. Collusion resistance between a malicious user and a semi-honest *Converter* $\mathcal{C}$, thanks to the *convertible tag* that is verified by each user after receiving their tags by the *Converter*. *Convertible tags* allow each user to verify whether or not $\mathcal{C}$ tried to forge a convertible tag after colluding with another user.
3. Collusion resistance between a malicious Aggregator $\mathcal{A}$ and a semi-honest *Converter* $\mathcal{C}$. In the case of users who do not act maliciously, meaning they have not been captured by an external adversary, who shares secret information with a malicious Aggregator or *Converter*, our protocol is resilient to collusions between a malicious Aggregator $\mathcal{A}$ and a semi-honest *Converter* $\mathcal{C}$.

As we extend the model for privacy preserving and unforgeable aggregation as presented in [38] and in section 2.1, with malicious users and extra parties (*Converter*) in the protocol, we also change the model of the scheme syntactically and we describe it as follows.

4.2 Collusion Resistant Aggregation Model

- $\mathsf{Setup}(1^\lambda)$: This is a probabilistic algorithm that on input the security parameter λ it outputs the public parameters pp and the secret key sk_A of the Aggregator.
- $\mathsf{UKeygen}(1^\lambda)\langle \mathcal{KD}, \mathbb{U}\rangle$: The key dealer $\mathcal{KD}$ runs this algorithm in order to distribute secret keys to each user for encryption. Moreover users choose uniformly at random their tag keys.
- $\mathsf{CKeygen}(1^\lambda)\langle \mathcal{KD}, \mathbb{U}, \mathcal{C}, \mathcal{DA}\rangle$: This key distribution algorithm runs between the users who blind their randomness from the $\mathsf{UKeygen}(1^\lambda)\langle \mathcal{KD}, \mathbb{U}\rangle$ algorithm, send that to the Converter $\mathcal{C}$, and the latter distributes the secret authentication tag key to the data analyzer $\mathcal{DA}$.
- $\mathsf{EncTag}(\mathsf{pp}, \mathsf{sk}_i, x_{i,t})$: Each user using its secret encryption key encrypts its individual data and sends the ciphertext $c_{i,t}$ to $\mathcal{A}$. Moreover using its secret tag key computes a metatag $\mathsf{mtag}_{i,t}$ and sends that to the *Converter* $\mathcal{C}$.
- $\mathsf{Convert}(\mathsf{pp}, r, \mathsf{mtag}_{i,t})$: $\mathcal{C}$ with the key r, and the metatag $\mathsf{mtag}_{i,t}$ computes the final tag $\mathsf{st}_{i,t}$ for user $\mathcal{U}_i$.
- $\mathsf{VTag}(\mathsf{pp}, \mathsf{sk}_i, \mathsf{st}_{i,t}, x_{i,t})$: Each user verifies the correctness of the final tag $\mathsf{st}_{i,t}$. *Convertible tags* prevent $\mathcal{C}$ to forge a user's tag using secret information from a colluding user.
- $\mathsf{Aggregate}(\mathsf{sk}_A, \{c_{i,t}\}, \{\mathsf{st}_{i,t}\})$: Aggregator $\mathcal{A}$ upon collecting the ciphertexts $\{c_{i,t}\}$ and the tags $\{\mathsf{st}_{i,t}\}$ decrypts with the secret key sk_A and learns the

result $\mathsf{sum_t} = \sum_{i=1}^{n} x_{i,t}$. Moreover, it computes a proof of correct computation σ_t and finally and forwards to the data analyzer $\mathcal{DA}$ $\mathsf{sum_t}, \sigma_t$.
– $\mathsf{Verify}(\mathsf{pp}, \mathsf{vk}, \mathsf{sum_t}, \sigma_t)$: The data analyzer $\mathcal{DA}$ verifies the correctness of computation for the $\mathsf{sum_t}$, using the proof σ_t and the secret verification key vk and the public parameters pp.

4.3 Security and Privacy Model

In this section we analyze the collusions resiliency property for aggregation protocols. We further formally define the security and the privacy properties.

4.3.1 Collusions and Trust model

In contrast with previous model and solution [38], our scheme fulfills its security guarantees under weakened assumptions. More specifically, collusions in between users and malicious parties are supported without sacrificing the security definition for unforgeability. Users $\mathbb{U} = \{\mathcal{U}\}_{i=1}^{n}$ in the scheme are unauthenticated and can act maliciously. Collusions can happen between a malicious user $\mathcal{U}_m$ and a colluding Aggregator $\mathcal{A}$ or a malicious *Converter* $\mathcal{C}$. Users share any secret information they know with the colluding members with the goal to forge other users' tag. Users are only trusted to submit correct data be it malicious or honest but curious. Collusions between $\mathcal{C}$ and $\mathcal{A}$ are also possible in case of users do not collude with $\mathcal{A}$ or $\mathcal{C}$ (cf. table 1). In case of colluding $\mathcal{A}$ and $\mathcal{C}$ and at least 1 malicious user, then the protocol does not provide aggregate unforgeability. We also assume the data analyzer $\mathcal{DA}$ to be a trustworthy party, who does not communicate with the users. We thus, omit it from the security model. We first describe the oracles an adversary $\mathcal{A}$ has access to when collusions between $\mathcal{U}, \mathcal{C}$ and $\mathcal{A}$ are possible:

– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$: On input a user identifier uid this oracle transmits to an adversary who impersonates a malicious Aggregator $\mathcal{A}$ the user's secret information $(\mathsf{ek_{uid}}, \mathsf{tk_{uid}}, \mathsf{r_{uid}}, w)$ after running the $\mathsf{UKeygen}(1^\lambda)$ and $\mathsf{CKeygen}(1^\lambda)$ algorithms.
– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$ On input user identifier uid this oracle runs the $\mathsf{UKeygen}(1^\lambda)$ and $\mathsf{CKeygen}(1^\lambda)$ algorithms and forwards to a malicious *Converter* $\mathcal{C}$ the user's secret information $(\mathsf{ek_{uid}}, \mathsf{tk_{uid}}, \mathsf{r_{uid}}, w)$.
– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$ In case of honest but curious users this oracle returns the secret key of $\mathcal{C}$ to an adversary $\mathcal{A}$.

User	$\mathsf{Coll}_{\mathcal{A},\mathcal{C}}$	$\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}$	$\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}$
HbC	✓	✓	✓
Malicious	✗	✓	✓

Table 1: Collusion model for Aggregate unforgeability. HbC denotes honest but curious users who follow the rules of the protocol but they may collude with $\mathcal{A}$ or $\mathcal{C}$.

4.3.2 Collusion Resistant Aggregate Unforgeability

The security of the scheme is modeled under the *collusion resistant aggregate unforgeability* ($\mathsf{CR-AU}$) security definition. An adversary $\mathcal{A}$ is able to obtain valid authentication tags for values of its choice by corrupting users. $\mathcal{A}$ also learns valid encryptions of its choice, and learns the final result over plaintext values $\sum_{i=1}^{n} x_{i,t}$. In the end we claim that an aggregation scheme is secure if a malicious Aggregator $\mathcal{A}$ cannot forge an aggregate tag for a time interval t^* such that for the underlying plaintexts it holds that $\sum_{i=1}^{n} x_{i,t^*} \neq \sum_{i=1}^{n} x_{i,t}$ for a set of users $\mathcal{U}_i \in \mathbb{S}$ that did not collude with the Aggregator or the *Converter*. We follow the security syntax as in [38] and we differentiate between:

– **Type-I** forgeries, in which $\mathcal{A}$ tries to forge for a time interval t^* in which she has not seen any tags from the users.
– **Type-II** forgeries for a time interval t, in which $\mathcal{A}$ has received valid tags for the users but $\mathsf{sum_t}^* \neq \sum x_{i,t}$.

However, in our model we allow a malicious Aggregator or Converter to collude with a user, in pursuance of forging another user's tag and convince the honest data analyzer $\mathcal{DA}$ for the correctness of computations given erroneous data inputs. An adversary during the $\mathsf{CR-AU}$ game has access to the following oracles:

– $\mathcal{O}^{\mathsf{Setup}}()$: This oracle when queried responds with the public parameters of the scheme pp and the secret key of the Aggregator sk_A.
– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$: On input a user identifier uid, this oracle when is queried by a malicious Aggregator $\mathcal{A}$ replies with the secret key of a user $\mathsf{sk_{uid}}$.
– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$: Upon receiving a user identifier uid the $\mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}$ oracle responds to a malicious *Converter* $\mathcal{C}$ with the secret key of a user $\mathsf{sk_{uid}}$.
– $\mathcal{O}^{\mathsf{Corr}_{\mathcal{A}}}()$: This oracles responds with the secret decryption key sk_A of the Aggregator.
– $\mathcal{O}^{\mathsf{Corr}_{\mathcal{C}}}()$: This oracles responds with the secret key of the *Converter* $\mathcal{C}$.
– $\mathcal{O}^{\mathsf{Corr}_{\mathcal{DA}}}()$: This oracles responds with the secret verification key vk of the data analyzer $\mathcal{DA}$.
– $\mathcal{O}_{\mathcal{A}}^{\mathsf{EncTag}}(t, \mathsf{uid}, x_{i,t})$: This is an oracle that replies with the encryption of the value $x_{i,t}$ using the secret

key of the user i after calling the $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(t, \mathsf{uid} = i \in \mathbb{U})$ or $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(t, \mathsf{uid} = i \in \mathbb{U})$ oracle. It also returns the metatag $\mathsf{mtag}_{i,t}$

- $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$: The $\mathcal{O}^{\mathsf{Mtag}}$ oracle on input a metatag $\mathsf{mtag}_{i,t}$ it converts it to the tag $\mathsf{st}_{i,t}$ after corrupting *Converter's* secret key with the $\mathcal{O}^{\mathsf{Corr}_{\mathcal{C}}}$ oracle.
- $\mathcal{O}_{\mathcal{A}}^{\mathsf{Aggregate}}(\{c_{i,t}\}_{i=1}^{n})$: This oracle simulates the behavior of the Aggregator $\mathcal{A}$ and when invoked with inputs the ciphertexts $\{c_{i,t}\}_{i=1}^{n}$, it gives as a response the sum $\sum_{i=1}^{n} x_{i,t}$, after calling the $\mathcal{O}^{\mathsf{Corr}_{\mathcal{A}}}$ oracle, in order to obtain the secret decryption key of the Aggregator sk_A.
- $\mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$: Upon receiving a tuple, containing a time interval t, a proof σ_{t} and and a result $\mathsf{sum}_{\mathsf{t}}$, the $\mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}$ oracle invokes the $\mathcal{O}^{\mathsf{Corr}_{\mathcal{DA}}}$ oracle and replies with $1 \iff \mathsf{sum}_{\mathsf{t}} = \sum_{i=1}^{n} x_{i,t}$, or $\perp$ otherwise.

We model the security definition of $\mathsf{CR} - \mathsf{AU}$, with two games: $\mathbf{Game}^{\mathsf{CR}-\mathsf{AU}-I}$ and $\mathbf{Game}^{\mathsf{CR}-\mathsf{AU}-II}$ respectively.

In $\mathbf{Game}^{\mathsf{CR}-\mathsf{AU}-I}$ users act maliciously and collusions between a user and a malicious Aggregator $\mathcal{A}$ or a $\mathcal{C}$ are allowed. During the learning phase of the game (cf. algorithm 1), $\mathcal{A}$ interacts with $\mathcal{O}^{\mathsf{Setup}}(), \mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U}), \mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U}), \mathcal{O}_{\mathcal{A}}^{\mathsf{EncTag}}(t, \mathsf{uid}, x_{i,t}), \mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t}), \mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$ oracles, in order to get the public parameters pp, the secret tag key of the user, allow the *Converter* to collude with a malicious user, the ciphertexts, the tags and the metatags of a user, respectively. Finally through $\mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$ $\mathcal{A}$ has access to the verification oracle. Note, that this oracle during the game makes sense since, our scheme operates in a symmetric setting, thus $\mathcal{A}$ cannot publicly verify. Finally $\mathcal{A}$ outputs a forgery for a time interval t^*. The forgery is successful if $\mathsf{Verify}(\mathsf{pp}, \mathsf{vk}, \mathsf{sum}_t^*, \mathsf{st}_t^*) = 1$ for a time interval t^* in which $\mathcal{A}$ did not query the $\mathcal{O}_{\mathsf{EncTag}}$ (**Type-I** forgery), or for t^* in which $\mathcal{A}$ called $\mathcal{O}_{\mathsf{EncTag}}$ (**Type-II** forgery) and none of users $\mathcal{U}_i \in \mathbb{S}$ collude with the Aggregator or the *Converter*.

Definition 1 $(\mathsf{CR} - \mathsf{AU} - I)$ An aggregation scheme is $\mathsf{CR} - \mathsf{AU} - I$ secure if any probabilistic polynomial time adversary $\mathcal{A}$ has negligible probability $\epsilon(\lambda)$ on the winning probabilities $\Pr[\mathcal{A}^{\mathsf{CR}-\mathsf{AU}-I}(\lambda)])$ of the game as describe in algorithms 1, 2: $\Pr[\mathcal{A}^{\mathsf{CR}-\mathsf{AU}-I}(\lambda)] \leq \epsilon(\lambda)$.

In $\mathbf{Game}^{\mathsf{CR}-\mathsf{AU}-II}$ users do not collude with $\mathcal{A}$ or $\mathcal{C}$ but collusions between $\mathcal{C}$ and $\mathcal{A}$ can occur. During the security game though, in the learning phase (cf. algorithm 3) $\mathcal{A}$ does not have access to the $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$ and $\mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$ oracles during which

1 : $(\mathsf{pp}, \mathsf{sk}_A) \leftarrow \mathcal{O}_{\mathsf{Setup}}(1^\lambda)$
2 : // $\mathcal{A}$ executes the following a polynomial number of times
3 : $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$
4 : $\mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$
5 : // $\mathcal{A}$ is allowed **to** call $\mathcal{O}_{\mathsf{EncTag}}$
6 : **for** *all users* $\in \mathcal{U}_i$ **do**
7 : $(c_{i,t}, \mathsf{st}_{i,t}) \leftarrow \mathcal{O}_{\mathsf{EncTag}}(t, \mathsf{uid}_i, x_{i,t})$
8 : $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$
9 : $\mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$

Fig. 1: Learning phase of the $\mathsf{CR} - \mathsf{AU} - I$ game

1 : $(t^*, \mathsf{sum}_{t^*}, \sigma_{t^*}) \leftarrow \mathcal{A}$

Fig. 2: Challenge phase of the $\mathsf{CR} - \mathsf{AU} - I$ game

1 : $(\mathsf{pp}, \mathsf{sk}_A) \leftarrow \mathcal{O}_{\mathsf{Setup}}(1^\lambda)$
2 : // $\mathcal{A}$ executes the following a polynomial number of times
3 : // $\mathcal{A}$ is allowed **to** call $\mathcal{O}_{\mathsf{EncTag}}$
4 : **for** *all users* $\in \mathcal{U}_i$
5 : $(c_{i,t}, \mathsf{st}_{i,t}) \leftarrow \mathcal{O}_{\mathsf{EncTag}}(t, \mathsf{uid}_i, x_{i,t})$
6 : $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$
7 : $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$
8 : $\mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$

Fig. 3: Learning phase of the $\mathsf{CR} - \mathsf{AU} - II$ game

1 : $(t^*, \mathsf{sum}_{t^*}, \sigma_{t^*}) \leftarrow \mathcal{A}$

Fig. 4: Challenge phase of the $\mathsf{CR} - \mathsf{AU} - II$ game

users share their secret keys with $\mathcal{A}$ and $\mathcal{C}$. However, $\mathcal{A}$ has access to $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$ oracle since Aggregator and *Converter* can collude. Similarly with $\mathbf{Game}^{\mathsf{CR}-\mathsf{AU}-I}$ $\mathcal{A}$ succeeds if it outputs during the challenge phase (cf. algorithm 4) either a **Type-I** or **Type-II** forgery.

Correspondingly for a $\mathsf{CR} - \mathsf{AU} - II$ scheme we define:

Definition 2 $(\mathsf{CR} - \mathsf{AU} - II)$ An aggregation scheme is $\mathsf{CR} - \mathsf{AU} - II$ secure if any probabilistic polynomial time adversary $\mathcal{A}$ has negligible probability $\epsilon(\lambda)$ on the winning probabilities $\Pr[\mathcal{A}^{\mathsf{CR}-\mathsf{AU}-II}(\lambda)])$ of the game as describe in algorithms 3, 4: $\Pr[\mathcal{A}^{\mathsf{CR}-\mathsf{AU}-II}(\lambda)] \leq \epsilon(\lambda)$.

$1:\ (\mathsf{pp}, \mathsf{sk}_{\mathsf{A}}, \mathsf{vk}) \leftarrow \mathcal{O}_{\mathsf{Setup}}(1^{\lambda})$
$2:\ \mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$
$3:\ \mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$
$4:\ //\ \mathcal{A}$ executes the following a polynomial number of times
$5:\ //\ \mathcal{A}$ is allowed to call $\mathcal{O}_{\mathsf{EncTag}}$
$6:\ \textbf{for}\ \ users\ \in \mathcal{U}_i$
$7:\ (c_{i,t}, \mathsf{st}_{i,t}) \leftarrow \mathcal{O}_{EncTag}(t, \mathsf{uid}_i, x_{i,t})$
$8:\ \mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$
$9:\ \mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$

Fig. 5: Learning phase of the Aggregator obliviousness game

4.3.3 Aggregator Obliviousness

The privacy guarantees of the scheme assure Aggregator obliviousness (AO) as has been first modeled by Shi *et al.* [45] and followed in subsequent work [34, 37, 38]. In a nutshell, a malicious Aggregator $\mathcal{A}$ or *Converter* $\mathcal{C}$ cannot compromise individual privacy. $\mathcal{A}$ is allowed to learn in cleartext the sum over users' data inputs. The privacy definition has been augmented in order to capture the extra functionality of unforgeability. As such, an adversary $\mathcal{A}$ is able to observe apart from ciphertexts, metatags and the final convertible tag. Notice that all protocols so far are collusion resistant for privacy, since the sharing of the secret key of a user cannot compromise others' privacy by an adversarial Aggregator. We clarify that the collusion resistance property makes sense to the unforgeability security property of the protocol.

We are focused on AO since the Aggregator learns most of the information during the protocol execution. It is the party, which in contrast with the security definition of $\mathsf{CR}-\mathsf{AU}-I$ and $\mathsf{CR}-\mathsf{AU}-II$, has access to all collusion oracles $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U}), \mathcal{O}^{\mathsf{Coll}_{\mathcal{C},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U}), \mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$ during the learning phase of the game in algorithm 5. At the challenge phase (cf. algorithm 6), $\mathcal{A}$ chooses a subset of users $\mathbb{S}^*$ that have not been corrupted and issues two time series data $\mathcal{X}^0_{t^*} = (\mathcal{U}_i, t^*, x^0_{i,t^*})_{\mathcal{U}_i \in \mathbb{S}^*}$ and $\mathcal{X}^1_{t^*} = (\mathcal{U}_i, t^*, x^1_{i,t^*})_{\mathcal{U}_i \in \mathbb{S}^*}$, such that $\sum_{\mathcal{U}_i \in \mathbb{S}^*} x^0_{i,t^*} = \sum_{\mathcal{U}_i \in \mathbb{S}^*} x^1_{i,t^*}$ for a time interval t^* and sends them to the $\mathcal{O}_{\mathsf{AO}}(\mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*})$ oracle.

$\mathcal{O}_{\mathsf{AO}}(\mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*})$ upon receiving the time series data flips a random coin $b \xleftarrow{\$} \{0,1\}$ and replies to $\mathcal{A}$ with the ciphertexts $\{c^b_{i,t}\}_{\mathcal{U}_i \in \mathbb{S}^*}$ the metatags $\{\mathsf{mtag}^b_{i,t}\}_{\mathcal{U}_i \in \mathbb{S}^*}$ and the tags $\{\mathsf{st}^b_{i,t}\}_{\mathcal{U}_i \in \mathbb{S}^*}$. $\mathcal{A}$ can adaptively call the $\mathcal{O}_{\mathcal{A}}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$ oracle after the challenge.

At the end of the game $\mathcal{A}$ outputs its guess b^*, and $\mathcal{A}$ wins the game $\iff b^* = b$.

$1:\ \mathcal{A} \rightarrow t^*, \mathbb{S}^*$
$2:\ \mathcal{A} \rightarrow \mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*}$
$3:\ (c^b_{i,t^*}, \mathsf{st}^b_{i,t^*})_{\mathcal{U}_i \in \mathbb{S}^*} \leftarrow \mathcal{O}_{\mathsf{AO}}(\mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*})$
$4:\ \mathcal{A} \rightarrow b^*$

Fig. 6: Challenge phase of the Aggregator obliviousness game

Definition 3 (Aggregator Obliviousness) Let $\Pr[\mathcal{A}^{\mathbf{AO}}]$ denote the probability that Aggregator $\mathcal{A}$ outputs $b^* = b$. Then an aggregation protocol is said to ensure Aggregator obliviousness if for any polynomially bounded Aggregator $\mathcal{A}$ the probability $\Pr[\mathcal{A}^{\mathbf{AO}}] \leqslant \frac{1}{2} + \epsilon(\lambda)$, where ϵ is a negligible function and λ is the security parameter.

5 Preliminaries

In this section we explain the basic building blocks and computation assumptions that are used in our proofs.

5.1 Bilinear maps

Let $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ be cyclic groups of large prime order p and g_1, g_2 generators of $\mathbb{G}_1, \mathbb{G}_2$ accordingly. We say that e is a bilinear map, if the following properties are satisfied:

1. *bilinearity*: $e(g_1^a, g_2^b) = e(g_1, g_2)^{ab}$, where $g_1, g_2 \in \mathbb{G}_1 \times \mathbb{G}_2$ and $a, b \in \mathbb{Z}_p$.
2. *Computability*: there exists an efficient algorithm that computes $e(g_1^a, g_2^b)$ where $g_1, g_2 \in \mathbb{G}_1 \times \mathbb{G}_2$ and $a, b \in \mathbb{Z}_p$.
3. *Non-degeneracy*: $e(g_1, g_2) \neq 1$.

5.2 Computational Assumptions

Definition 4 (Bilinear Computational Diffie-Hellman (BCDH) Assumption)

Let $e(\mathbb{G}_1 \times \mathbb{G}_2) \rightarrow \mathbb{G}_T$ be a bilinear pairing, g a generator of $\mathbb{G}_1$ and g_2 a generator of $\mathbb{G}_2$ and p the order of $\mathbb{G}_1, \mathbb{G}_2$ and $\mathbb{G}_T$. Given $U = (g, g^a, g^b, g^c) \in \mathbb{G}_1$ and $V = (g_2, g_2^a, g_2^b) \in \mathbb{G}_2$ for random $a, b, c \in \mathbb{Z}_p$ we say that BCDH holds if the probabilities of a probabilistic polynomial time adversary $\mathcal{A}$ to compute $W = e(g_1, g_2)^{abc}$ are negligible on input the security parameter λ: $\Pr[W \leftarrow \mathcal{A}(U, V)]$.

Definition 5 (eXternal Diffie-Hellman (XDH) Assumption)

Let $e(\mathbb{G}_1 \times \mathbb{G}_2) \rightarrow \mathbb{G}_T$ be a bilinear pairing, g a generator of $\mathbb{G}_1$ and g_2 a generator of $\mathbb{G}_2$ and p the order

of $\mathbb{G}_1, \mathbb{G}_2$ and $\mathbb{G}_T$. We say that XDH holds if the probabilities of a probabilistic polynomial time adversary $\mathcal{A}$ to solve DDH and DL in $\mathbb{G}_1$ are negligible on input the security parameter λ.

Definition 6 (Fixed Argument Pairing Inversion I ($\mathsf{FAPI-I}$) Assumption) [29]

Let $e(\mathbb{G}_1 \times \mathbb{G}_2) \rightarrow \mathbb{G}_T$ be a bilinear pairing, $d_1 \in \mathbb{G}_1, d_2 \in \mathbb{G}_2$ and $e(d_1, d_2) = z \in \mathbb{G}_T$ We say that $\mathsf{FAPI-I}$ holds if the probabilities of a probabilistic polynomial time adversary who gets as input z and d_1 to output d_2 are negligible on input the security parameter λ: $\mathcal{A}$ $\Pr[d_2 \leftarrow \mathcal{A}(d_1, z)]$

5.3 Zero-Knowledge Proofs

Definition 7 During an interactive protocol $\langle P(x), V(x)\rangle$ between a probabilistic polynomial time prover P and a probabilistic polynomial time verifier V, the latter outputs `accept` as long as it accepts as correct the claim of the prover that $x \in \mathcal{L}$, for input x and a language $\mathcal{L}$. The interactive protocol is a zero knowledge proof system if the following properties are met:

- **Completeness**: For an honest prover P and V, who interact a common input x, then if the statement x is correct the verifier V is convinced with high probability:

$$\forall x \in \mathcal{L}, \Pr[\langle P(x), V(x)\rangle = \mathtt{accept}] \geq 1 - \mathsf{neg}(\lambda)$$

- **Soundness**: For any cheating prover P^*, the probability of an honest V to accept as correct a statement x which does not belong in $\mathcal{L}$ is negligible:

$$\forall P^*, \forall x \notin \mathcal{L}, \Pr[\langle P^*(x), V(x)\rangle = \mathtt{accept}] \leq \mathsf{neg}(\lambda)$$

- **Zero-knowledge**: The view of a cheating verifier V^* during the interactive protocol with the prover P does not reveal any further information. This is modeled with a simulator $\mathcal{S}$, who produces indistinguishable transcripts of the protocol by interacting with V^*:

$$\forall V^* \exists \mathcal{S} : \forall x \in \mathcal{L}, \mathsf{VIEW}_{P,V^*(x)} \approx \mathcal{S}(x)$$

.

In figure 7, we show how to construct a non-interactive zero knowledge proof for Pedersen style commitments $h^a g^b \bmod p, h = H(t), a, b \leftarrow \mathbb{Z}_q$, with the Fiat-Shamir heuristic [27], for a hash function $H : \{0,1\}^* \rightarrow \mathbb{Z}_p$, which is modeled as a random oracle in the zero-knowledge proof and the simulator $\mathcal{S}$ controls it.

Completeness follows by inspection. A honest verifier V validates $c^d y = c^{H(h,g,y,c)} y = [h^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i H(h,g,y,c)} y = h^{aH(h,g,y,c)} g^{bH(h,g,y,c)} y = h^{aH(h,g,y,c)} g^{bH(h,g,y,c)} h^{r_1} g^{r_2} = h^{aH(h,g,y,c)+r_1} g^{bH(h,g,y,c)+r_2} = h^{w_1} g^{w_2}$.

Soundness. We assume a malicious Prover P^* who convinces with non-negligible probability for some malformed $\mathsf{mtag}^1_{i,t}$. That is, P^* can convince an honest V for the same commit value c but on different values. This implies that given $(y, d, w_1, w_2), (y, d', w_1', w_2') \Rightarrow h^{w_1} g^{w_2} = y c^d, h^{w_1'} g^{w_2'} = y c^{d'}$. Thus, is true $h^{w_1} g^{w_2} = h^{w_1'} g^{w_2'} y^{d-d'}$. We conclude $a = \frac{w_1 - w_1'}{d-d'}, b = \frac{w_2 - w_2'}{d-d'}$. We can extract the discrete logs of h^a and g^b which contradicts the DL assumption.

Zero Knowledge. We assume that in the real world both the verifier and the prover can learn the evaluation of H(x) by querying the oracle. We use a simulator $\mathcal{S}$ who controls the random oracle H in the ideal world. $\mathcal{S}$ uses a lookup table to respond with the same value on equivalent inputs. Since $\mathcal{S}$ does not communicate with the prover it simulates the transcripts of the interaction as follows: $\mathcal{S}$ chooses $d, w_1, w_2 \xleftarrow{\$} \mathbb{Z}_p$, sets $y = \frac{h^{w_1} g^{w_2}}{c^d}$ and $H(y) = d$. Thus the view of the verifier in real world is indistinguishable with the simulation.

6 Protocol

In order to guarantee AO our protocol employs Shi *et al.* scheme [45]. For completeness we briefly describe their encryption scheme.

6.1 Shi-Chan-Rieffel-Chow-Song Scheme

- $\mathsf{Setup}(1^\lambda)$: On input the security parameter λ this probabilistic algorithm outputs a cryptographic secure hash function $H : \{0,1\}^* \rightarrow \mathbb{G}_1$, for a group $\mathbb{G}_1$ of large prime order p. Through a secure channel the trusted key dealer $\mathcal{KD}$ distributes to each user a secret encryption key $\mathsf{ek}_i \in \mathbb{Z}_p$, which is chosen uniformly at random. $\mathcal{KD}$ also forwards to the $\mathcal{A}$ the secret decryption key $\mathsf{sk}_\mathsf{A} = \sum_{i=1}^n \mathsf{ek}_i$.
- $\mathsf{Encrypt}(\mathsf{ek}_i, x_{i,t})$: To encrypt data value $x_{i,t}$ at time interval t with secret key ek_i, user $\mathcal{U}_i$ computes the ciphertext $c_{i,t} = H(t)^{\mathsf{ek}_i} g_1^{x_{i,t}} \in \mathbb{G}_1$.
- $\mathsf{Aggregate}(\{c_{i,t}\}_{\mathcal{U}_i \in \mathbb{U}}, \{\mathsf{st}_{i,t}\}_{\mathcal{U}_i \in \mathbb{U}}, \mathsf{sk}_\mathsf{A})$: Upon receiving all the ciphertexts $\{c_{i,t}\}_{i=1}^n$, the Aggregator computes: $V_t = (\prod_{i=1}^n c_{i,t}) H(t)^{-\mathsf{sk}_A} = H(t)^{\sum_{i=1}^n \mathsf{ek}_i} g_1^{\sum_{i=1}^n x_{i,t}} H(t)^{-\sum_{i=1}^n \mathsf{ek}_i} = g_1^{\sum_{i=1}^n x_{i,t}} \in \mathbb{G}_1$. $\mathcal{A}$ then learns the sum $\mathsf{sum}_\mathsf{t} = \sum_{i=1}^n x_{i,t} \in \mathbb{Z}_p$ by computing the discrete logarithm of V_t on the

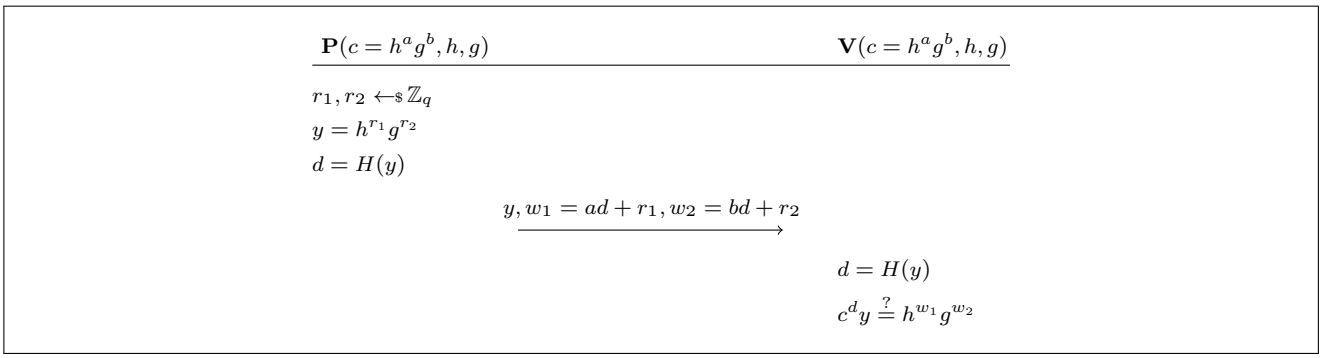

Fig. 7: $ZKP\{(a,b)|c = h^a g^b = \mathsf{mtag}^1_{i,t}\}$

base g_1. The sum computation is correct as long as $\sum_{i=1}^{n} x_{i,t} < p$.

6.2 Collusion resistant aggregation I (CRA-I)

In order to communicate the ideas of the protocol in a clear way we first define a protocol that is collusion resistant between colluding users and a malicious Aggregator $\mathcal{A}$.

- $\mathsf{Setup}_I(1^\lambda)$: On input the security parameter λ this probabilistic algorithm defines a cryptographic secure hash function $H : \{0,1\}^* \rightarrow \mathbb{G}_1$, a bilinear pairing $e : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$ of prime order p with generator g. Finally it outputs the public parameters $\mathsf{pp} = (H, e, g, g_2)$. It also calls the $\mathsf{Setup}(1^\lambda)$ algorithm of the Shi *et al.* scheme and outputs the secret key of the Aggregator sk_A
- $\mathsf{UKeygen}_I(1^\lambda)\langle \mathcal{KD}, \mathbb{U}\rangle$: Each user independently chooses uniformly random tag keys tk_i and r_i. Through a secure channel each $\mathcal{U}_i$ forwards r_i to the key dealer $\mathcal{KD}$, who computes $\sum_{i=1}^{n} \mathsf{r}_i$.
- $\mathsf{CKeygen}_I(1^\lambda)\langle \mathcal{KD}, \mathbb{U}\rangle$: The key dealer chooses uniformly at random a key $r \in \mathbb{Z}_p$ and a random generator $w \in \mathbb{G}_2$. It distributes through a secure channel r to the *Converter* $\mathcal{C}$. It also sends to the $\mathcal{DA}$ the secret verification key $\mathsf{vk} = (w, r, \sum_{i=1}^{n} \mathsf{r}_i)$. Moreover it forwards w to each user. Then the key dealer $\mathcal{KD}$ goes off-line.
- $\mathsf{EncTag}_I(\mathsf{pp}, \mathsf{sk_i}, \mathsf{x_{i,t}})$: This deterministic algorithm takes as input the secret key of each user $\mathsf{sk_i} = (\mathsf{r_i}, \mathsf{w}, \mathsf{tk_i}, \mathsf{ek_i})$ and the private values $x_{i,t}$ and outputs the metatag:

$$\mathsf{mtag}_{i,t} = (\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t}) = ([H(t)^{\mathsf{r_i}} g^{x_{i,t}}]^{\mathsf{tk_i}}, w^{\frac{1}{\mathsf{tk_i}}})$$

Moreover, users encrypt their data using the encryption key ek_i, with the encryption scheme of Shi *et al.* [45] as already presented in 6.1. Finally, $\mathcal{U}_i$ forwards $c_{i,t}$ to the Aggregator $\mathcal{A}$ and the metatag $\mathsf{mtag}_{i,t}$ to the Converter.

- $\mathsf{Convert}_I(\mathsf{pp}, r, \mathsf{mtag}_{i,t})$: The *Converter* runs this algorithm in order to "unify" all the tags under the same key. It allows the homomorphic operations on the tags. The algorithm takes as input the public parameters pp, the key r, and metatag $\mathsf{mtag}_{i,t}$ and outputs the tag $\mathsf{st}_{i,t}$ as follows:

$$\begin{aligned}\mathsf{st}_{i,t} &= e(\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t})^r = e([H(t)^{\mathsf{r_i}} g^{x_{i,t}}]^{\mathsf{tk_i}}, w^{\frac{1}{\mathsf{tk_i}}})^r \\ &= e(H(t)^{\mathsf{r_i}\ \mathsf{tk_i}}, w^{\frac{1}{\mathsf{tk}_i}})^r e(g^{x_{i,t}\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk_i}}})^r \\ &= e(H(t)^{\mathsf{r_i}}, w)^r e(g^{x_{i,t}}, w)^r\end{aligned}$$

- $\mathsf{Aggregate}_I(\mathsf{sk}_A, \{c_{i,t}\}, \{\mathsf{st}_{i,t}\})$: The Aggregator $\mathcal{A}$ after collecting all the ciphertexts $c_{i,t}$ for the users $\mathbb{U}$ decrypts with the secret key sk_A and learns the sum $\sum_{i=}^{n} x_{i,t}$. For the decryption algorithm $\mathcal{A}$ uses the decryption algorithm as in Shi *et al.* scheme [45]. Moreover, $\mathcal{A}$ computes a proof of correct computation by aggregating the tags $\mathsf{st}_{i,t}$ as follows:

$$\begin{aligned}\sigma_{\mathsf{t}} &= \prod_{i=1}^{n} \mathsf{st}_{i,t} = \prod_{i=1}^{n} e(\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t})^r \\ &= \prod_{i=1}^{n} e(H(t)^{\mathsf{r_i}}, w)^r e(g^{x_{i,t}}, w)^r \\ &= \prod_{i=1}^{n} e(H(t)^{\mathsf{r_i}}, w)^r \prod_{i=1}^{n} e(g^{x_{i,t}}, w)^r \\ &= e(H(t), w)^{r \sum_{i=1}^{n} \mathsf{r_i}} e(g, w)^{r \sum_{i=1}^{n} x_{i,t}}\end{aligned}$$

Finally $\mathcal{A}$ returns to the honest verifier the result $\mathsf{sum_t} = \sum_{i=}^{n} x_{i,t}$ and the proof $\sigma_{\mathsf{t}} = e(H(t), w)^{r \sum_{i=1}^{n} \mathsf{r_i}} e(g, w)^{r \sum_{i=1}^{n} x_{i,t}}$

- $\mathsf{Verify}_I(\mathsf{pp}, \mathsf{vk}, \mathsf{sum_t}, \sigma_{\mathsf{t}})$: The data analyzer $\mathcal{DA}$, who acts as honest verifier verifies the correctness of the sum computation by employing its verification key $\mathsf{vk} = (\mathsf{vk}_1 = w, \mathsf{vk}_2 = r, \mathsf{vk}_3 = \sum_{i=1}^{n} \mathsf{r}_i)$. $\mathcal{DA}$

verifies by checking if the following equation holds:

$$e(H(t), \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{sum}_\mathsf{t}} \stackrel{?}{=} \sigma_\mathsf{t}$$

Thanks to the bilinearity of the pairings the correctness of the verification procedure is assured. Indeed:

$$e(H(t), \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{sum}_\mathsf{t}} =$$
$$e(H(t), w)^{r\sum_{i=1}^n \mathsf{r}_i} e(g, w)^{r\sum_{i=1}^n x_{i,t}} = \sigma_\mathsf{t}$$

6.3 Collusion resistant aggregation II (CRA-II)

We now present an extension of the previous scheme in order mitigate collusions between users and a malicious $\mathcal{A}$ and between users and malicious $\mathcal{C}$, meaning that a user can collude at the same time with $\mathcal{A}$ and $\mathcal{C}$. First we define a simple attack on the previous scheme:

Attack on CRA-I scheme A colluding user $\mathcal{U}_c$ shares with the *Converter* his secret tag key tk_i and the shared common key between all users w. $\mathcal{C}$ can forge a valid tag $\mathsf{tag}_{i,t}$ for a user $\mathcal{U}_i$ as follows: $\mathsf{tag}'_{i,t} = \mathsf{tag}_{i,t} \cdot e(g^{x'_{i,t}}, w) = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}+x'_{i,t}}, w)^r$, which is a valid forge for the value $x_{i,t} + x'_{i,t}$.

The core idea to mitigate these type of attacks is to enforce the *Converter* $\mathcal{C}$ to re-randomize the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}$ with the randomness r, such that $\mathcal{C}$ replies to $\mathcal{U}_i$ with the final tag $\mathsf{st}_{i,t} = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$ along with the randomized metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{r\mathsf{tk}_i}$. Finally the user recomputes the final tag from the randomized metatag and validates whether the final tag has been forged. As such, in case of collusions between a malicious user and a malicious $\mathcal{C}$, the latter can forge the final tag, but the user can detect it, thanks to the unforgeability of the metatag $\mathsf{mtag}_{i,t}$. Notice that this mitigation assumes honest behavior of the users. Due to this, malicious users without colluding with the *Converter* $\mathcal{C}$ can tamper metatags in order to enforce *Converter* $\mathcal{C}$ to reply with w^r. This secret information allows the Aggregator to collude with a user in order to forge a tag of a benign user and finally to authenticate the sum over non-genuine data. We circumvent this with a zero-knowledge proof that assures the valid form of a metatag. Thus users can behave arbitrary without affecting the security of the scheme. We describe the entire protocol for collusion resistant aggregation against $\mathcal{C}$ and $\mathcal{A}$:

- $\mathsf{Setup}_{II}(1^\lambda)$: This algorithm calls the $\mathsf{Setup}_I(1^\lambda)$ algorithm and outputs the public parameters $\mathsf{pp} = (H, e, g, g_2)$ and the secret key of the Aggregator sk_A
- $\mathsf{UKeygen}_{II}(1^\lambda)\langle\mathcal{KD}, \mathbb{U}\rangle$: $\mathsf{UKeygen}_{II}(1^\lambda)$ invokes the $\mathsf{UKeygen}_I(1^\lambda)$ algorithm during which each user independently chooses uniformly random tag keys tk_i and r_i. Moreover users transmit r_i, through a secure channel to the key dealer who computes $\sum_{i=1}^n \mathsf{r}_i$.
- $\mathsf{CKeygen}_{II}(1^\lambda)\langle\mathcal{KD}, \mathbb{U}, \mathcal{C}, \mathcal{DA}\rangle$: This algorithm calls the $\mathsf{CKeygen}_I(1^\lambda)\langle\mathcal{KD}, \mathbb{U}, \mathcal{C}, \mathcal{DA}\rangle$, in which the key dealer outputs the secret verification key $\mathsf{vk} = (w, r, \sum_{i=1}^n \mathsf{r}_i)$, chooses uniformly at random a key $r \in \mathbb{Z}_p$ and a random generator $w \in \mathbb{G}_2$. It distributes through a secure channel r to the *Converter* $\mathcal{C}$. It also sends w to each user, and forwards to the $\mathcal{DA}$ the secret verification key $\mathsf{vk} = (w, r, \sum_{i=1}^n \mathsf{r}_i)$. Finally the key dealer $\mathcal{KD}$ goes off-line.
- $\mathsf{EncTag}_{II}(\mathsf{pp}, \mathsf{sk}_i, x_{i,t})$: $\mathsf{EncTag}_{II}(\mathsf{pp}, \mathsf{sk}_i, x_{i,t})$ calls $\mathsf{EncTag}_I(\mathsf{pp}, \mathsf{sk}_i, x_{i,t})$ and operates similarly. It outputs for each user $\mathcal{U}_i$ the ciphertext $c_{i,t}$ and the metatag :

$$\mathsf{mtag}_{i,t} = (\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t}) = ([H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}})$$

which is forwarded to the *Converter* $\mathcal{C}$.
- $\mathsf{Convert}_{II}(\mathsf{pp}, r, \mathsf{mtag}_{i,t})$: Upon receiving the metatag $\mathsf{mtag}_{i,t} = (\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t}) = ([H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}})$, $\mathcal{C}$ acts a verifier and user $\mathcal{U}_i$ as a prover and proves to $\mathcal{C}$ in zero knowledge:

$$ZKP\{a = \mathsf{r}_i\mathsf{tk}_i, b = x_{i,t}\mathsf{tk}_i | c = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i} = \mathsf{mtag}^1_{i,t}\}$$

If the proof fails the protocol aborts. Otherwise $\mathcal{C}$ uses its secret key r to compute the final tag as follows:

$$\begin{aligned}\mathsf{st}^1_{i,t} &= e(\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t})^r \\ &= e([H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}})^r \\ &= e(H(t)^{\mathsf{r}_i \cdot \mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}})^r e(g^{x \cdot \mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}})^r \\ &= e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r\end{aligned}$$

$$\mathsf{st}^2_{i,t} = (\mathsf{mtag}^1_{i,t})^r = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{r\mathsf{tk}_i}$$

[1]

Finally $\mathcal{C}$ sends to $\mathcal{U}_i$ the final tag $\mathsf{st}_{i,t} = (\mathsf{st}^1_{i,t}, \mathsf{st}^2_{i,t})$.
- $\mathsf{VTag}_{II}(\mathsf{pp}, \mathsf{sk}_i, \mathsf{st}_{i,t}, x_{i,t})$: Each user verifies the correctness of the final tag as follows:

$$e(\mathsf{st}^2_{i,t}, w)^{\frac{1}{\mathsf{tk}_i}} \stackrel{?}{=} \mathsf{st}^1_{i,t}$$

The correctness of the equation holds since:

$$\begin{aligned}e(\mathsf{st}^2_{i,t}, w)^{\frac{1}{\mathsf{tk}_i}} &= e([H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{r\mathsf{tk}_i}, w)^{\frac{1}{\mathsf{tk}_i}} \\ &= e(H(t)^{\mathsf{r}_i} g^{x_{i,t}}, w)^{\frac{r\mathsf{tk}_i}{\mathsf{tk}_i}} e(H(t)^{\mathsf{r}_i} g^{x_{i,t}}, w)^r \\ &= \mathsf{st}^1_{i,t}\end{aligned}$$

[1] Gray background denotes the different crypto machinery needed to prevent collusions between users and $\mathcal{C}$.

At this point if the equation is not true the user $\mathcal{U}_i$ halts the execution of the protocol and it infers that $\mathcal{C}$ forged the tag $\mathsf{st}_{i,t}$. Otherwise it continues by sending the final tag $\mathsf{st}_{i,t} = \mathsf{st}^1_{i,t}$ to the Aggregator $\mathcal{A}$.

– $\mathsf{Aggregate}_{II}(\mathsf{sk}_A, \{c_{i,t}\}, \{\mathsf{st}_{i,t}\})$: This algorithm calls $\mathsf{Aggregate}_I(\mathsf{sk}_A, \{c_{i,t}\}, \{\mathsf{st}_{i,t}\})$, which consecutively decrypts with the secret key sk_A and $\mathcal{A}$ learns $\mathsf{sum_t} = \sum_{i=1}^{n} x_{i,t}$. Moreover, it computes a proof of correct computation σ_t and finally and forwards the result $\mathsf{sum_t} = \sum_{i=}^{n} x_{i,t}$ and the proof $\sigma_\mathsf{t} = \prod_{i=1}^{n} \mathsf{st}_{i,t} = e(H(t), w)^{r \sum_{i=1}^{n} \mathsf{r_i}} e(g, w)^{r \sum_{i=1}^{n} x_{i,t}}$ to the data analyzer $\mathcal{DA}$.

– $\mathsf{Verify}_{II}(\mathsf{pp}, \mathsf{vk}, \mathsf{sum_t}, \sigma_\mathsf{t})$: The $\mathsf{Verify}_{II}(\mathsf{pp}, \mathsf{vk}, \mathsf{sum_t}, \sigma_\mathsf{t})$ algorithm invokes $\mathsf{Verify}_I(\mathsf{pp}, \mathsf{vk}, \mathsf{sum_t}, \sigma_\mathsf{t})$ and verifies the correctness of the sum computation by checking :

$$e(H(t), \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum_t}} \stackrel{?}{=} \sigma_\mathsf{t}$$

7 Security Analysis

In this section we give evidence for the security of the scheme, following the security definitions in section 4.3. We start our analysis with privacy and we prove the *Aggregator unforgeability* privacy property. Notice that be it **CRA-I** or **CRA-II** the privacy guarantee is not affected as with the encryption scheme of Shi *et al.* [45] in case of corrupted users, thanks to the trusted key dealer, who distributes individual secret keys to each user. As such, we assume a trusted key distribution phase before the key dealer $\mathcal{KD}$ goes off-line.

7.1 Aggregator Obliviousness

Theorem 1 *The **CRA-I** and **CRA-II** schemes provide Aggregator Obliviousness under the DDH assumption in $\mathbb{G}_1$ in the random oracle mode.*

Proof We assume an adversary $\mathcal{A}$ who breaks with non-negligible probability the AO privacy definition for *Aggregator obliviousness*. We will show in our proof how a probabilistic polynomial time adversary $\mathcal{B}$ invokes $\mathcal{A}$ as a subroutine in order to break the *Aggregator obliviousness* definition as defined in the scheme of Shi *et al.* [45]. We will refer to this scheme as private streaming aggregation (PSA). Adversary $\mathcal{B}$ has access to $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Setup}}$, $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Corrupt}}$, $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Encrypt}}$, and $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{AO}}$ oracles with the challenger, when she tries to break AO in PSA. The $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Setup}}$ oracle gives the public parameters and the secret keys to the users and the Aggregator. The $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Corrupt}}$ oracle on input a user id uid returns the secret encryption key sk_i of a corrupted user. The $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Encrypt}}$ oracle on input a data input $x_{i,t}$ returns the encryption $c_{i,t}$ under the encryption algorithm of [45]. The $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{AO}}$ oracle during the challenge phase with $\mathcal{B}$ flips a random coin $b \xleftarrow{\$} \{0,1\}$ and responds with the encryption of the time series $\mathcal{X}^b_t = \{x_{i,t}\}$.

Algorithm $\mathcal{B}$ simulates as a challenger the oracles $\mathcal{A}$ has access to during the **Learning phase** as follows:

– $\mathcal{O}_{\mathsf{Setup}}(1^\lambda)$: Whenever $\mathcal{A}$ calls the $\mathcal{O}_{\mathsf{Setup}}(1^\lambda)$ oracle, $\mathcal{B}$ calls the $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Setup}}$ oracle, which responds to $\mathcal{B}$ with a hash function $H : \{0,1\}^* \to \mathbb{G}_1$, a generator g of the group $\mathbb{G}_1$ of safe prime order p, and the Aggregator's secret key $\mathsf{sk}_A = \sum_{i=1}^{n} \mathsf{ek_i}$. Moreover, $\mathcal{B}$ chooses the parameters of a bilinear pairing $bp = (e, g_1, g_2, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T)$. Uniformly at random it selects secret keys $r, \{\mathsf{r_i}\}_{i=1}^n, \{\gamma_i\}_{i=1}^n \in \mathbb{Z}_p, w \in \mathbb{G}_2$. Finally $\mathcal{B}$ replies to $\mathcal{A}$ with H, g, p, sk_A.

– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$: When $\mathcal{A}$ invokes this oracle then $\mathcal{B}$ calls the $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Corrupt}}$ oracle and transmits to $\mathcal{A}$ the secret encryption key ek_c of a corrupted user $\mathcal{U}_c \in \mathbb{S}$ and its secret tag key $\mathsf{r}_i, \gamma_i, w$.

– $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$: The collusion between the *Converter* and $\mathcal{A}$ are simulated by the $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{C}}}(\mathsf{uid} = i \in \mathbb{U})$ oracle. $\mathcal{B}$ replies on these calls with the secret key r.

– $\mathcal{O}_{\mathsf{EncTag}(t,\mathsf{uid},x_{i,t})}$: Upon call on the $\mathcal{O}_{\mathsf{EncTag}(t,\mathsf{uid},x_{i,t})}$ oracle, $\mathcal{B}$ invokes the $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{Encrypt}}$ with input $(t, \mathsf{uid}, x_{i,t})$, which in turns reply to $\mathcal{B}$ with the encryption $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}}$ of $x_{i,t}$. $\mathcal{B}$ also computes $\mathsf{mtag}_{i,t} = [c^{\mathsf{r}_i}_{i,t}, w^{\frac{1}{\mathsf{r}_i}}] = [(H(t)^{\mathsf{ek}_i} g^{x_{i,t}})^{\mathsf{r}_i}, w^{\frac{1}{\mathsf{r}_i}}]$. Notice that $\mathsf{mtag}_{i,t}$ is indistinguishable from the real one if we interchange the randomness and set $\mathsf{r}_i = \mathsf{ek}_i$ and $\mathsf{tk}_i = \mathsf{r}_i$, for uniformly random keys $\mathsf{ek}_i, \mathsf{r}_i$. $\mathcal{B}$. Finally $\mathcal{B}$ replies to $\mathcal{A}$ with $(c_{i,t}, \mathsf{mtag}_i, t)$.

– $\mathcal{O}^{\mathsf{Mtag}}_{\mathcal{A}}(\mathsf{mtag}_{i,t})$: $\mathcal{A}$ calls this oracle in order to learn the final tag of each user $\mathsf{st}_{i,t}$. $\mathcal{B}$ computes the final tag as $\mathsf{st}_{i,t} = e(\mathsf{mtag}^1_{i,t}, \mathsf{mtag}^2_{i,t}) = e((H(t)^{\mathsf{ek}_i} g^{x_{i,t}})^{\mathsf{r}_i}, w^{\frac{1}{\mathsf{r}_i}})^r = e(H(t)^{\mathsf{ek}_i}, w)^r e(g^{x_{i,t}}, w)^r$. Under the verification key $\mathsf{vk} = (w, r, \mathsf{sk}_A)$ the aggregation of the tags $\prod \mathsf{st}_{i,t}$ can be correctly verified, upon calling the $\mathcal{O}^{\mathsf{Verify}}_{\mathcal{A}}(t, \sigma_\mathsf{t}, \mathsf{sum_t})$ oracle.

– $\mathcal{O}^{\mathsf{Verify}}_{\mathcal{A}}(t, \sigma_\mathsf{t}, \mathsf{sum_t})$: $\mathcal{A}$ can query this oracle to learn the result of verification. We assume an honest verifier and this oracle makes sense, since we are in a symmetric verifications setting with a secret key. $\mathcal{B}$ returns the result of the verification since it knows the secret verification key $\mathsf{vk} = (\mathsf{vk}_1, \mathsf{vk}_2, \mathsf{vk}_3) = (w, r, \mathsf{sk}_A,)$:

$$e(H(t), \mathsf{vk}_1)^{\mathsf{vk}_2, \mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum_t}} \stackrel{?}{=} \sigma_\mathsf{t}$$

When the learning phase is over, then $\mathcal{A}$ during the **Challenge phase**, chooses a set of users $\in \mathbb{S}^*$, that

have not been corrupted during the **Learning phase** and chooses two time series $\mathcal{X}_0^* = (\mathcal{U}_i, t^*, x^0_{i,t^*})_{\mathcal{U}_i \in \mathbb{S}^*}$ and $\mathcal{X}_1^* = (\mathcal{U}_i, t^*, x^1_{i,t^*})_{\mathcal{U}_i \in \mathbb{S}^*}$ such that $\sum x^0_{i,t^*} = \sum x^1_{i,t^*}$ for a time interval t^* in which $\mathcal{A}$ did not query neither the $\mathcal{O}_{\mathsf{EncTag}}$ nor the $\mathcal{O}^{\mathsf{Mtag}}$ oracle and sends them to $\mathcal{O}_{\mathsf{AO}}(\mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*})$ oracle.

To simulate $\mathcal{O}_{\mathsf{AO}}(\mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*})$ $\mathcal{B}$ queries the $\mathcal{O}^{\mathsf{PSA}}_{\mathsf{AO}}$ oracle with input $\mathcal{X}^0_{t^*}, \mathcal{X}^1_{t^*}$, which in turns flips a random coin $b \xleftarrow{\$} \{0,1\}$ and responds to $\mathcal{B}$ with the ciphertexts $\{c^b_{i,t^*}\}_{\mathcal{U}_i \in \mathbb{S}^*}$. $\mathcal{B}$ also computes the final tags:

$$\begin{aligned}\mathsf{st}^b_{i,t^*} &= e((H(t^*)^{\mathsf{ek}_i} g^{x^b_{i,t^*}})^{\mathsf{r}_i}, w^{\frac{1}{\mathsf{r}_i}})^r \\ &= e(H(t^*)^{\mathsf{ek}_i}, w)^r e(g^{x^b_{i,t^*}}, w^r)\end{aligned}$$

Finally $\mathcal{B}$ forwards $\{c^b_{i,t^*}, \mathsf{st}^b_{i,t^*}\}$ to $\mathcal{A}$. The tag st^b_{i,t^*} simulates perfectly the final tag of a user and the aggregation of the tags for the computation of the final proof σ_{t} correctly verifies the sum under the secret verification key $\mathsf{vk} = (\mathsf{vk}_1, \mathsf{vk}_2, \mathsf{vk}_3) = (w, r, \mathsf{sk}_A)$:

$$\begin{aligned}\prod_{i=1}^{n} \mathsf{st}^b_{i,t^*} &= \prod_{i=1}^{n} e(H(t^*)^{\mathsf{ek}_i}, w)^r e(g^{x^b_{i,t^*}}, w^r) \\ &= \sigma_{\mathsf{t}} = e(H(t), \mathsf{vk}_1)^{\mathsf{vk}_2, \mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}}\end{aligned}$$

If $\mathcal{A}$ has non-negligible advantage ϵ to correctly guess the bit b^* for the bit b, then $\mathcal{B}$ will break the AO game in the PSA scheme with non-negligible advantage ϵ. This contradicts the DDH assumption since the security of PSA is reduced to the DDH assumption. As such our scheme assures AO in the random oracle model under the XDH assumption, which assumes the intractability of DDH in $\mathbb{G}_1$.

7.2 Aggregate unforgeability

Type-I Forgeries

Theorem 2 *An adversary A who colludes with a user $\mathcal{U}_c$ in the **CRA-I** scheme has negligible probability on forging a **Type-I** $\mathsf{CR}-\mathsf{AU}-I$ forgery, under the* BCDH *assumption in the random oracle mode.*

We prove theorem 2 in three steps. First we prove the security of a base scheme (BaseLine) without any collusions in between a user and any other party. To model this scheme, an adversary $\mathcal{A}$ plays the game as described in algorithms 1 and 2 without access to the corruption oracles $\mathcal{O}^{\mathsf{Corr}_{\mathcal{A}}}, \mathcal{O}^{\mathsf{Corr}_{\mathcal{C}}}$ and $\mathcal{O}^{\mathsf{Corr}_{\mathcal{DA}}}$. For the sake of clarity we call the security definition of *aggregate unforgeability* in the BaseLine scheme as BAU and the corresponding game $\mathbf{Game}^{\mathsf{BAU}}$. Then we show that a **Type-I** forgery in the **CRA-I** can be transformed to a **Type-I** forgery in the BaseLine scheme and finally that a **Type-I** forgery in the **CRA-II** scheme can be transformed to a **Type-I** forgery in the BaseLine scheme, as well.

Lemma 1 *The baseline scheme guarantees* aggregate unforgeability *for **Type-I** forgeries under the* BCDH *assumption in the random oracle model.*

Proof We will show how an adversary $\mathcal{B}$ injects the challenge of the BCDH assumption into the game that adversary $\mathcal{A}$ plays. During the setup phase $\mathcal{B}$ receives the challenge $(g, g_2, g^a, g^b, g^c, g_2^a, g_2^b)$ from $\mathcal{O}^{\mathsf{BCDH}}_{\mathsf{Setup}}$ oracle and is asked to output $e(g, g_2)^{abc}$. $\mathcal{B}$ simulates the Challenger when $\mathcal{A}$ plays the $\mathbf{Game}^{\mathsf{BAU}}$ game as follows:

$\mathcal{B}$ first chooses uniformly at random secret keys $\mathsf{w}, \mathsf{r}, \{\mathsf{r}_i, \mathsf{ek}_i, \mathsf{tk}_i\}_{i=1}^n$

Learning phase

- $\mathcal{O}_{\mathsf{Setup}}$: Whenever $\mathcal{A}$ calls this oracle, $\mathcal{B}$ returns the public parameters $\mathsf{pp} = (H, e, g, g_2)$ for a hash function $H : \{0,1\}^* \to \mathbb{G}_1$, bilinear pairing e, generators g, g_2 for $\mathbb{G}_1, \mathbb{G}_2$ and the secret key of the Aggregator $\mathsf{sk}_A = \sum_{i=1}^n \mathsf{ek}_i$. $\mathcal{B}$ also sets as the secret verification key $\mathsf{vk} = (g_2^a, r, g^{b\sum_{i=1}^n \mathsf{r}_i})$ and does not share this information.
- $\mathcal{A}$ can query the random oracle H for a time interval t. In order to respond to the queries $\mathcal{B}$ constructs a list $\mathtt{H_L}\langle t : v_t, \mathsf{coin}(t), H(t)\rangle$ and responds to $\mathcal{A}$ as follows:
  - If H has been queried before at the time interval t, $\mathcal{B}$ fetches the tuple $\mathtt{H_L}(t)$ and replies to $\mathcal{A}$ with $H(t)$.
  - If t is fresh then $\mathcal{B}$ selects uniformly at random $\phi_t \in \mathbb{Z}_p$ and flips a random $\mathsf{coin}(t)$. With probability p $\mathsf{coin}(t) = 0$ and $\mathcal{B}$ appends to $\mathtt{H_L}(t) = g^{\phi_t}$. Otherwise with probability $1-p$ when $\mathsf{coin}(t) = 1$ then $\mathcal{B}$ sets $\mathtt{H_L}(t) = g^{c\phi_t}$. Finally $\mathcal{B}$ sends $\mathtt{H_L}(t)$ to $\mathcal{A}$.
  - Whenever $\mathcal{A}$ calls the $\mathcal{O}_{\mathsf{EncTag}(t,\mathsf{uid},x_{i,t})}$ oracle, $\mathcal{B}$ constructs a tuple $\mathtt{ET}\langle t, \mathsf{uid}_i, x_{i,t}, \mathsf{st}_{i,t}\rangle$. We differentiate three cases:
    1. If at time interval t, $\mathcal{O}_{\mathsf{EncTag}(t,\mathsf{uid},x_{i,t})}$ has not been queried before, then $\mathcal{B}$ calls the simulated random oracle for time interval t and gets the response $H(t)$. If $\mathtt{coin}(t) = 1$ then $\mathcal{B}$ halts the simulation. Otherwise it computes the ciphertext with the secret encryption key ek_i as $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}} = g^{\phi_t \mathsf{ek}_i} g^{x_{i,t}}$. Finally $\mathcal{B}$ computes the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}$ and forwards $c_{i,t}, \mathsf{mtag}_{i,t}$ to $\mathcal{A}$. It also updates $\mathtt{ET}$ list with the tuple: $\langle t, \mathsf{uid}_i, x_{i,t}, \mathsf{st}_{i,t}\rangle$ and sets $\Sigma_t = \Sigma_t + x_{i,t}$.

2. If there exists uid in the list $\mathtt{ET}$ for time interval t, then $\mathcal{B}$ fetches this tuple and forwards $c_{i,t}, \mathsf{st}_{i,t}$ to $\mathcal{A}$.
3. Else $\mathcal{B}$ fetches the corresponding tuple from the $\mathtt{H_L}$ list. If $\mathtt{coin}(t) = 1$ then $\mathcal{B}$ halts the simulation. Otherwise it computes the ciphertext with the secret encryption key ek_i as $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}} = g^{\phi_t \mathsf{ek}_i} g^{x_{i,t}}$. Finally $\mathcal{B}$ computes the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}$ and forwards $c_{i,t}, \mathsf{mtag}_{i,t}$ to $\mathcal{A}$. It also updates $\mathtt{ET}$ list with the tuple: $\langle t, \mathsf{uid}_i, x_{i,t}, \mathsf{st}_{i,t} \rangle$ and sets $\Sigma_t = \Sigma_t + x_{i,t}$.

– When $\mathcal{A}$ calls the $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$ oracle, $\mathcal{B}$ calls the simulated random oracle to get $H(t)$. If $\mathtt{coin}(t) = 1$ then $\mathcal{B}$ halts, otherwise it forwards to $\mathcal{A}$ $\mathsf{st}_{i,t} = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$.

Challenge phase At the challenge phase $\mathcal{A}$ outputs a forgery $\mathsf{sum_t}^*, \sigma_{\mathsf{t}}{}^*$ for a time interval t^*. $\mathcal{B}$ fetches the tuple $\mathtt{H_L}(t^*)$ and:

– If $\mathtt{coin}(t^*) = 0$, then it aborts.
– Otherwise it solves the BCDH assumption by computing:

$$
\begin{aligned}
I &= \frac{(\sigma_{\mathsf{t}}{}^*)}{e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}{}^*}} \\
&= \frac{e(H(t^*), \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}{}^*}}{e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}{}^*}} \\
&= e(H(t^*), \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{vk}_3} = e(g^{c\phi_{t^*}}, g_2^a)^{rb \sum_{i=1}^n \mathsf{r}_i}
\end{aligned}
$$

Finally it outputs $I^{\frac{1}{\phi_{t^*} r \sum_{i=1}^n \mathsf{r}_i}} = e(g, g_2)^{abc}$, which is the solution to the BCDH problem.

The probabilities of $\mathcal{B}$ to not abort are $p^2(1-p)^{q_h}$ for q_h queries to the random oracle. So assuming $\mathcal{A}$ forges a **Type-I** forgery with some non-negligible probability $\epsilon'(\lambda)$, then $\Pr[\mathcal{B}^{\mathsf{BCDH}}] = p^2(1-p)^{q_h}\epsilon'(\lambda)$. As such we ended up in a contradiction assuming the hardness of the BCDH assumption and $\Pr[\mathcal{A}^{\mathsf{BAU}}] = \epsilon(\lambda)$ for some negligible function ϵ on input of the security parameter λ.

Lemma 2 *Let $\mathcal{A}$ be a probabilistic polynomial time adversary who colludes with a user $\mathcal{U}_c$ in the **CRA-I** scheme and outputs a **Type-I** forgery with non-negligible probability. Then, there is an adversary $\mathcal{B}$ that outputs a **Type-I** forgery for the* $\mathtt{BaseLine}$ *scheme with non-negligible probability.*

Proof $\mathcal{B}$ calls the $\mathcal{O}_{CRA-I}^{\mathsf{Setup}}$ oracle which returns the public parameters $\mathsf{pp} = (H, e, g, g_2)$ and the secret key of the Aggregator sk_A. $\mathcal{B}$ relays this information to $\mathcal{A}$. Whenever $\mathcal{A}$ calls the $\mathcal{O}_{\mathcal{A}}^{\mathsf{EncTag}}(t, \mathsf{uid}, x_{i,t})$ oracle, then $\mathcal{B}$ in turn forwards the query to the $\mathcal{O}_{\mathcal{A}}^{\mathsf{EncTag}}(t, \mathsf{uid}, x_{i,t})$ oracle of the **CRA-I** game, which replies with $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}}, \mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}$. Similarly $\mathcal{B}$ relays the queries to the $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$ and forwards the response $= \mathsf{st}_{i,t} = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$ back to $\mathcal{A}$. $\mathcal{B}$ responds to the queries for the $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(t, \mathsf{uid} = i \in \mathbb{U})$ oracle, with $(\mathsf{r}_i, \mathsf{ek}_i, \mathsf{tk}_i, w)$. Note that for honest but curious users, $\mathcal{A}$ only learns $H(t)^{\mathsf{ek}_i} g^{x_{i,t}}, [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}, e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$ by knowing w. Thus the secret value $x_{i,t}$ and the secret keys of the user are computationally hidden. At this point the view of $\mathcal{A}$ is consistent with the real protocol and thus does not abort the game. Eventually $\mathcal{A}$ outputs a forgery $\sigma_{\mathsf{t}}{}^*$. $\mathcal{B}$ also outputs $\sigma_{\mathsf{t}}{}^*$ as a valid forgery.

Lemma 3 *Let $\mathcal{C}$ be a probabilistic polynomial time adversary who colludes with a user $\mathcal{U}_c$ in the **CRA-II** scheme and outputs a **Type-I** forgery with non-negligible probability. Then, there is an adversary $\mathcal{B}$ that outputs a **Type-I** forgery for the* $\mathtt{BaseLine}$ *scheme with non-negligible probability.*

Proof The proof proceeds accordingly with the previous proof for lemma 2. $\mathcal{B}$ relays queries to $\mathcal{O}_{CRA-II}^*$ oracles, coming from $\mathcal{C}$. When $\mathcal{C}$ corrupts a user $\mathcal{U}_i \in \mathbb{S}$ then $\mathcal{B}$ forwards to $\mathcal{C}$ the secret keys $(\mathsf{r}_i, \mathsf{ek}_i, \mathsf{tk}_i, w)$. Finally the view of adversary $\mathcal{C}$ is identical with the real game without being able to distinguish since $H(t)^{\mathsf{ek}_i} g^{x_{i,t}}, [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}, e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$ computationally hide the secret value $x_{i,t}$ and $(\mathsf{r}_i, \mathsf{ek}_i, \mathsf{tk}_i,)$ keys from uncorrupted users by an adversary $\mathcal{C}$ knowing the secret secret key r and secret keys of corrupted users.

With lemmas 1, 2, 3 we conclude the proof of theorem 2.

Theorem 3 *An adversary A has negligible probability on forging a **Type-I** $\mathsf{CR-AU-}II$ forgery, under the* BCDH *assumption in the random oracle mode.*

Notice the a $\mathsf{CR-AU-}II$ forgery entails collusions between a *Converter* and an Aggregator, by revealing r to the latter. Thus the proofs proceeds as with lemma 1, with the difference that $\mathcal{A}$ during the learning phase calls the $\mathcal{O}^{\mathsf{Corr}_C}$ oracle and $\mathcal{B}$ forwards to $\mathcal{A}$ the secret key r.

Proof $\mathcal{B}$ first chooses uniformly at random secret keys $\mathsf{w}, \mathsf{r}, \{\mathsf{r}_i, \mathsf{ek}_i, \mathsf{tk}_i\}_{i=1}^n$

Learning phase

– $\mathcal{O}_{\mathsf{Setup}}$: Whenever $\mathcal{A}$ calls this oracle, $\mathcal{B}$ returns the public parameters $\mathsf{pp} = (H, e, g, g_2)$ for a hash function $H : \{0,1\}^* \to \mathbb{G}_1$, bilinear pairing e, generators g, g_2 for $\mathbb{G}_1, \mathbb{G}_2$ and the secret key of the Aggregator $\mathsf{sk}_A = \sum_{i=1}^{n} \mathsf{ek}_i$. $\mathcal{B}$ also sets as the secret verification key $\mathsf{vk} = (g_2^a, r, g^{b\sum_{i=1}^{n} \mathsf{r}_i})$ and does not share this information.
– $\mathcal{A}$ can query the random oracle H for a time interval t. In order to respond to the queries $\mathcal{B}$ constructs a list $\mathtt{H_L}\langle t : v_t, \mathsf{coin}(t), H(t)\rangle$ and responds to $\mathcal{A}$ as follows:
  – If H has been queried before at the time interval t, $\mathcal{B}$ fetches the tuple $\mathtt{H_L}(t)$ and replies to $\mathcal{A}$ with $H(t)$.
  – If t is fresh then $\mathcal{B}$ selects uniformly at random $\phi_t \in \mathbb{Z}_p$ and flips a random $\mathsf{coin}(t)$. With probability p $\mathsf{coin}(t) = 0$ and $\mathcal{B}$ appends to $\mathtt{H_L}(t) = g^{\phi_t}$. Otherwise with probability $1-p$ when $\mathsf{coin}(t) = 1$ then $\mathcal{B}$ sets $\mathtt{H_L}(t) = g^{c\phi_t}$. Finally $\mathcal{B}$ sends $\mathtt{H_L}(t)$ to $\mathcal{A}$.
  – Whenever $\mathcal{A}$ calls the $\mathcal{O}_{\mathsf{EncTag}(t,\mathsf{uid},x_{i,t})}$ oracle, $\mathcal{B}$ constructs a tuple $\mathtt{ET}\langle t, \mathsf{uid}_i, x_{i,t}, \mathsf{st}_{i,t}\rangle$. We differentiate three cases:
    1. If at time interval t, $\mathcal{O}_{\mathsf{EncTag}(t,\mathsf{uid},x_{i,t})}$ has not been queried before, then $\mathcal{B}$ calls the simulated random oracle for time interval t and gets the response $H(t)$. If $\mathsf{coin}(t) = 1$ then $\mathcal{B}$ halts the simulation. Otherwise it computes the ciphertext with the secret encryption key ek_i as $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}} = g^{\phi_t \mathsf{ek}_i} g^{x_{i,t}}$. Finally $\mathcal{B}$ computes the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}$ along with a zero knowledge proof of $zkp\{a = \mathsf{r}_i\mathsf{tk}_i, b = x_{i,t}\mathsf{tk}_i | c = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i} = \mathsf{mtag}^1_{i,t}\}$ and forwards $c_{i,t}, \mathsf{mtag}_{i,t}, zkp$ to $\mathcal{A}$. It also updates $\mathtt{ET}$ list with the tuple: $\langle t, \mathsf{uid}_i, x_{i,t}, \mathsf{st}_{i,t}\rangle$ and sets $\Sigma_t = \Sigma_t + x_{i,t}$.
    2. If there exists uid in the list $\mathtt{ET}$ for time interval t, then $\mathcal{B}$ fetches this tuple and forwards $c_{i,t}, \mathsf{st}_{i,t}$ to $\mathcal{A}$ with a zero knowledge proof of $zkp\{a = \mathsf{r}_i\mathsf{tk}_i, b = x_{i,t}\mathsf{tk}_i | c = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i} = \mathsf{mtag}^1_{i,t}\}$.
    3. Else $\mathcal{B}$ fetches the corresponding tuple from the $\mathtt{H_L}$ list. If $\mathsf{coin}(t) = 1$ then $\mathcal{B}$ halts the simulation. Otherwise it computes the ciphertext with the secret encryption key ek_i as $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}} = g^{\phi_t \mathsf{ek}_i} g^{x_{i,t}}$. Finally $\mathcal{B}$ computes the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}$ and forwards $c_{i,t}, \mathsf{mtag}_{i,t}$ to $\mathcal{A}$ along with a with a zero knowledge proof of $zkp\{a = \mathsf{r}_i\mathsf{tk}_i, b = x_{i,t}\mathsf{tk}_i | c = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i} = \mathsf{mtag}^1_{i,t}\}$. It also updates $\mathtt{ET}$ list with the tuple: $\langle t, \mathsf{uid}_i, x_{i,t}, \mathsf{st}_{i,t}\rangle$ and sets $\Sigma_t = \Sigma_t + x_{i,t}$.
– $\mathcal{B}$ forwards to $\mathcal{A}$ the secret key r while invoking the $\mathcal{O}^{\mathsf{Corr}c}$ oracle.
– When $\mathcal{A}$ calls the $\mathcal{O}^{\mathsf{Mtag}}_{\mathcal{A}}(\mathsf{mtag}_{i,t})$ oracle, $\mathcal{B}$ calls the simulated random oracle to get $H(t)$. If $\mathsf{coin}(t) = 1$ then $\mathcal{B}$ halts, otherwise it forwards to $\mathcal{A}$ $\mathsf{st}_{i,t} = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$.

Challenge phase At the challenge phase $\mathcal{A}$ outputs a forgery $\mathsf{sum_t}^*, \sigma_{\mathsf{t}}{}^*$ for a time interval t^*. $\mathcal{B}$ fetches the tuple $\mathtt{H_L}(t^*)$ and:

– If $\mathsf{coin}(t^*) = 0$, or $zkp = \perp$, then it aborts.
– Otherwise it solves the BCDH assumption by computing:

$$I = \frac{(\sigma_{\mathsf{t}}{}^*)}{e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}{}^*}}$$
$$= \frac{e(H(t^*), \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}{}^*}}{e(g, \mathsf{vk}_1)^{\mathsf{vk}_2 \mathsf{sum}_{\mathsf{t}}{}^*}}$$
$$= e(H(t^*), \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{vk}_3} = e(g^{c\phi_{t^*}}, g_2^a)^{rb\sum_{i=1}^{n} \mathsf{r}_i}$$

Finally it outputs $I^{\frac{1}{\phi_{t^*} r \sum_{i=1}^{n} \mathsf{r}_i}} = e(g, g_2)^{abc}$, which is the solution to the BCDH problem.

Similarly with Lemma 1 the probabilities of $\mathcal{B}$ to not abort are $p^2(1-p)^{q_h}$ for q_h queries to the random oracle. $\mathcal{A}$ outputs a **Type-I** $\mathsf{CR} - \mathsf{AU} - II$ forgery with some non-negligible probability $\epsilon'(\lambda)$, then $\Pr[\mathcal{B}^{\mathsf{BCDH}}] = p^2(1-p)^{q_h}\epsilon'(\lambda)$, which is a contradiction assuming the hardness of BCDH assumption and concludes the proof.

Type-II Forgeries

For the proof of the theorem 4 we first introduce the following assumption:

Definition 8 (Dual Fixed Argument Pairing Inversion I ($\mathsf{DFAPI} - \mathsf{I}$) Assumption)

Let $e(\mathbb{G}_1 \times \mathbb{G}_2) \to \mathbb{G}_T$ be a bilinear pairing, $c_1, d_1 \in \mathbb{G}_1, c_2, d_2 \in \mathbb{G}_2$ and $e(c_1, c_2) = z_1 \in \mathbb{G}_T$, $e(d_1, d_2) = z_2 \in \mathbb{G}_T$. We say that $\mathsf{FAPI} - \mathsf{I}$ holds if the probabilities of a probabilistic polynomial time adversary $\mathcal{A}$ $\Pr[d_2 \leftarrow \mathcal{A}(d_1, z_1 \cdot z_2)]$ are negligible on input the security parameter λ.

For the proof of the aforementioned assumption we will show how a probabilistic polynomial time adversary $\mathcal{A}$ who has non-negligible probabilities on the $\mathsf{DFAPI} - \mathsf{I}$ assumption, can be used by a probabilistic polynomial time adversary $\mathcal{B}$ to break the $\mathsf{FAPI} - \mathsf{I}$ assumption with non negligible probabilities. We denote by $\mathcal{O}^{\mathsf{FAPI}-\mathsf{I}}$ the oracle of the $\mathsf{FAPI} - \mathsf{I}$ assumption, which outputs the challenge to an adversary and by $\mathcal{O}^{\mathsf{DFAPI}-\mathsf{I}}$ the oracle of the $\mathsf{DFAPI} - \mathsf{I}$ assumption.

Proof $\mathcal{B}$ queries the $\mathcal{O}^{\mathsf{FAPI-I}}$ oracle and gets back $(e, d_1, z = e(d_1, d_2))$. When $\mathcal{A}$ asks the $\mathcal{O}^{\mathsf{DFAPI-I}}$ for the public parameters of the scheme then $\mathcal{B}$ computes $y = e(r_1, r_2)e(d_1, d_2)$ by choosing $r_1 \in \mathbb{G}_1, r_2 \in \mathbb{G}_2$ and forwards it to $\mathcal{A}$. Assuming $\mathcal{A}$ breaks the $\mathsf{DFAPI-I}$ assumption with non-negligible probability ϵ by outputting d_2, then $\mathcal{B}$ outputs d_2 as a solution to $\mathsf{FAPI-I}$ with equivalent non-negligible probability ϵ.

Theorem 4 *An adversary $\mathcal{A}$ who colludes with a user $\mathcal{U}_c$ in the **CRA-II** scheme has negligible probability on forging a **Type-II** $\mathsf{CR-AU-}I,II$ forgery, under the $\mathsf{DFAPI-I}$ assumption in the standard model.*

Proof We show the interaction of $\mathcal{A}$ with the oracles, who eventually in order to provide a valid forgery has to solve the $\mathsf{DFAPI-I}$ assumption with non-negligible probability.

- $\mathtt{flag} = 0$.
- $\mathcal{O}_{\mathsf{Setup}}$: Whenever $\mathcal{A}$ calls this oracle, it receives the public parameters $\mathsf{pp} = (H, e, g, g_2)$ for a hash function $H : \{0,1\}^* \rightarrow \mathbb{G}_1$, bilinear pairing e, generators g, g_2 for $\mathbb{G}_1, \mathbb{G}_2$ and the secret key of the Aggregator $\mathsf{sk}_A = \sum_{i=1}^{n} \mathsf{ek}_i$. This oracle also sets as the secret verification key $\mathsf{vk} = (g_2^a, r, g^{b\sum_{i=1}^{n} \mathsf{r}_i})$ and does not share this information.
- $\mathcal{O}^{\mathsf{Corr}c}$: This oracle forwards to $\mathcal{A}$ the secret key r while invoking the $\mathcal{O}^{\mathsf{Corr}c}$ oracle and sets $\mathtt{flag} = 1$, to indicate a $\mathsf{CR-AU-}II$ forgery.
- $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$: If $\mathtt{flag} == 0$ then $\mathcal{O}^{\mathsf{Coll}_{\mathcal{A},\mathcal{U}_m}}(\mathsf{uid} = i \in \mathbb{U})$transmits the secret keys $\mathsf{ek}_i, \mathsf{tk}_i, r_i, w$. Otherwise it returns $\mathtt{null}$ to $\mathcal{A}$.
- $\mathcal{O}_{\mathcal{A}}^{\mathsf{EncTag}}(t, \mathsf{uid}, x_{i,t})$: $\mathcal{A}$ receives the ciphertext with the secret encryption key ek_i as $c_{i,t} = H(t)^{\mathsf{ek}_i} g^{x_{i,t}}$, the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i}, w^{\frac{1}{\mathsf{tk}_i}}$ and a zero knowledge proof of $zkp\{a = \mathsf{r}_i\mathsf{tk}_i, b = x_{i,t}\mathsf{tk}_i | c = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^{\mathsf{tk}_i} = \mathsf{mtag}_{i,t}^1\}$
- $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$: When $\mathcal{A}$ calls the $\mathcal{O}_{\mathcal{A}}^{\mathsf{Mtag}}(\mathsf{mtag}_{i,t})$ oracle, it gets $\mathsf{st}_{i,t} = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$.

Eventually $\mathcal{A}$ outputs a forgery for a time interval t, $\mathsf{st}_{i,t}$ for a $\mathsf{sum}_{\mathsf{t}} = s$. If $zkp = \perp$ then $\mathcal{A}$ aborts. In case of a $\mathsf{CR-AU-}I$ forgery then $\mathtt{flag} = 0$ and $\mathcal{A}$ learns $\mathsf{ek}_i, \mathsf{tk}_i, r_i, w$ for some users. In order the forgery to be valid and be accepted by the $\mathcal{O}^{\mathsf{Verify}}(t, \sigma_{\mathsf{t}}, \mathsf{sum}_{\mathsf{t}})$ oracle it ought to have the following form: $\mathsf{st}_{i,t} = e(H(t)^{\sum \mathsf{r}_i}, w)^r e(g^s, w)^r = e(H(t)^{\sum \mathsf{r}_i}, w)^r e(g^s, w^r)$ for $s = \mathsf{sum}_{\mathsf{t}}$. As such, in order $\mathcal{A}$ to compute $\mathsf{st}_{i,t}$ needs to extract r which is coupled with each tag $\prod \mathsf{st}_{i,t} = e(H(t)^{\sum \mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$, which is an instance of a $\mathsf{FAPI-I}$ assumption with $c_1 = H(t)^{\sum \mathsf{r}_i}, c_2 = w^r, d_1 = g^s, d_2 = w^r$ and its hardness is proved in theorem 8.

Similarly, in case the forgery is of type $\mathsf{CR-AU-}II$ then $\mathtt{flag} = 1$ and $\mathcal{A}$ knows r, s and $H(t)$ from $\mathsf{st}_{i,t}$. In order the forgery to be valid $\mathcal{A}$ needs to extract w from $\mathsf{st}_{i,t} = e(H(t)^{\sum \mathsf{r}_i}, w)^r e(g^s, w)^r$. From the bilinearity the equation can be expanded as $\mathsf{st}_{i,t} = e(H(t)^{\sum \mathsf{r}_i}, w)^r e(g^{rs}, w)$. The latter is an instance of a $\mathsf{FAPI-I}$ assumption with $c_1 = H(t)^{\sum \mathsf{r}_i}, c_2 = w^r, d_1 = g^{rs}, d_2 = w$.

8 Cost Analysis

8.1 Overhead

We perform a theoretical evaluation of the scheme with respect to the cardinality of operations that have to be performed by each party during the protocol execution for collusion resistant unforgeability. The results are depicted in table 2. Notice that we omit from the analysis the computational costs for encryption per user and decryption time for the Aggregator, since our goal is to show the extra cost for collusion resistant unforgeability. That is the costs to compute tags and metatags, convert them, per user verify them, aggregate them at the Aggregator's side and compute/verify the proof of correct computation.

At each time interval for the computation of the metatag $\mathsf{mtag}_{i,t} = [H(t)^{\mathsf{r}_i} g^{x_{i,t}}]^r, w^{\frac{1}{\mathsf{tk}_i}}$, $\mathcal{U}_i$ is committed to two exponentiations in $\mathbb{G}_1$ and one exponentiation in $\mathbb{G}_2$. Afterwards, in order to validate the final tag, users check if the following equation holds: $e(\mathsf{st}_{i,t}^2, w)^{\frac{1}{\mathsf{tk}_i}} \stackrel{?}{=} \mathsf{st}_{i,t}^2$, by performing one exponentiation in $\mathbb{G}_T$ and one bilinear pairing operation. Moreover for the zero knowledge proof its user computes 2 exponentiations in $\mathbb{G}_1$. The *Converter*, in order to convert the metatag $\mathsf{mtag}_{i,t}$ to the final tag $= \mathsf{st}_{i,t} = e(H(t)^{\mathsf{r}_i}, w)^r e(g^{x_{i,t}}, w)^r$, is committed to one bilinear pairing computation and one exponentiation in $\mathbb{G}_T$. For the verification of the zero-knowledge proof the *Converter* performs 3 exponentiations in $\mathbb{G}_1$. The Aggregator computes the proof with $n-1$ multiplications in $\mathbb{G}_T$: $\sigma_{\mathsf{t}} = \prod_{i=1}^{n} \mathsf{st}_{i,t} = e(H(t), w)^{r\sum_{i=1}^{n} \mathsf{r}_i} e(g, w)^{r\sum_{i=1}^{n} x_{i,t}}$ and the data analyzer verifies with two multiplications in $\mathbb{G}_T$, two exponentiations in $\mathbb{G}_T$ and two bilinear pairing evaluations: $e(H(t), \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{vk}_3} e(g, \mathsf{vk}_1)^{\mathsf{vk}_2\mathsf{sum}_{\mathsf{t}}} \stackrel{?}{=} \sigma_{\mathsf{t}}$. Notice, that the protocol achieves constant verification time, which does not dependent on the number of users that participate in the protocol with their values.

We also performed a real world prototype implementation for the verification process on a machine running Ubuntu 14.04 with kernel version 3.19.0-29. The machine has 8MB RAM memory and is equipped with an INTEL Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz processor with 4 cores. For our prototype implementation we used python version 3 and charm cryp-

Participant	Computation	Communication
User	$4E^{\mathbb{G}_1} + 1E^{\mathbb{G}_2} + 1E^{\mathbb{G}_T} + 1H + 1I + 1P$	$5l + l_T$
Converter	$4E^{\mathbb{G}_1} + 1E^{\mathbb{G}_T} + 1H + 2M + 1P$	$l + l_T$
Aggregator	$(n-1)M^{\mathbb{G}_T}$	$l + l_T$
Data Analyzer	$2E^{\mathbb{G}_T} + 1H + 2M^{\mathbb{Z}_p} + 1M^{\mathbb{G}_T} + 2P$	-

Table 2: Performance of tag computation and metatag computation, proof construction and verification operations. l denotes the bit-size of the prime number p and l_T the bit-size of elements in $\mathbb{G}_T$. E: exponentiation, I: inverse, H: hash, M: multiplication, P: pairing

tographic framework[1]. Charm supports an abstract layer for basic cryptographic primitives and its core system for the mathematical operations is implemented in ANSI C, yielding substantial efficiency results. Since out protocol is based on bilinear pairings we run our benchmarks with different implemented elliptic curves (cf. table 3). We measured the exact time of computing the metatags, the final tag and the verification of the final tag from the user side. Secondly, we measured the computational overhead at the converter side for transforming the metatags of each user. Finally, the time for verification is computed, for different curve types. The implementation results are obtained on average after running them for 1000 times. The cost for the Converter to convert each tag remains small compared with the computation of each tag, thanks to the simplicity of the Convert operations. The exact computational cost for verifying is also significantly low and it is constant, independent on the number of the users.

Table 3: Benchmark results

Operation	Curve Type		
	MNT159	MNT201	MNT224
User : Tag	**78** ms	**102** ms	**111** ms
Converter : Convert	**57** ms	**97** ms	**110** ms
DA : Verify	**80** ms	**103** ms	**112** ms

8.2 Discussion

Converter. At this point, we argue about the need of the Converter $\mathcal{C}$, who acts as an honest-but-curious party. The Aggregator $\mathcal{A}$ receives all the final tags $\{e(H(t)^{r_i} g^{x_{i,t}}, w)^r\}_{i=1}^n$. Its product constitutes the proof $\mathcal{A}$ sends to the Data Analyzer: $\sigma_t = \prod_{i=1}^n \mathsf{st}_{i,t} = e(H(t), w)^{r \sum_{i=1}^n r_i} e(g, w)^{r \sum_{i=1}^n x_{i,t}}$. Notice that the final sum is accumulated in the exponent of the second bilinear map, blinded with r. As long as r is known by the users then the protocol does not provide collusion unforgeability as with protocol [38]. Thus, a party must provide this secret information to the users in such a way that users do not learn about it. On the other hand the final tag must include extra randomness by each user such that individual privacy is preserved. The Converter acts as this honest-but-curious intermediate party who injects this extra randomness to the tags. However, it is not trusted as it can compromise individual privacy. The $\mathsf{Convert}_{II}$ algorithm in our protocol provides the privacy guarantees for individual privacy.

Fault Tolerance. If we would like to propose a collusion resistant aggregation protocol with fault tolerance then it seems that the extra property does not come for free. Naive solutions assume the online existence of a fully trusted key-dealer, which distributes new decryption key to the Aggregator each time a user fails to submit its encrypted data [21]. The online presence of a key-dealer is far from being realistic. Moreover, it becomes a single point of failure for external adversaries.

To mitigate the existence of a fully trusted key dealer during the protocol execution [16] proposes fault tolerant techniques by organizing users in a tree node and recursively aggregate in a bottom-up fashion till the root, which acts as the Aggregator. When a user fails to report its private input then the protocol tries to estimate the sum of contiguous intervals of users, which are modeled as the leaves of the tree. This technique even though it allows the execution of the protocol in case of faults, it introduces considerable noise, which might not be acceptable in use case scenarios with precision in mind as ours. Think for instance a billing protocol. Users will not accept to participate if the protocol amends positive additive noise to the calculation of their meterings. To reduce the error some papers [20, 30, 39] relies on the communication of the users without any intermediate party. That assumes the apriory knowledge of all users in the network, something which is also not that realistic in a smart grid scenario for instance as smart meters do not know the presence of others. Moreover in terms of network bandwidth the protocols with multi-user communication may be prohibitive.

All the aforementioned solutions do not include verifiability in their work, as such they cannot really be compared with ours. The only solution that meets the requirement for fault tolerance with communication efficiency and precise aggregation results without errors

is the work in [37], which does not provide verifiability. In that work users independently choose their randomness, which is blinded before sent to an honest-but-curious Collector. The latter computes the blinded decryption key and forwards it to the Aggregator. Fault tolerance comes for free as each time a user fails to encrypt its data, the Collector computes the blinded aggregation of the remaining users and forwards it to the Aggregator. The authors assume the non-collusion of the Collector and the Aggregator. Collusion between users and any party does not really make sense as each user has its own key and the security functionality assures individual privacy. As such sharing a secret key with the Aggregator, will not compromise the privacy of another user (as long as the colluding users remain less than $n-1$). The solution employs an algebraic group of $\mathbb{Z}_{N^2}$, taking advantage Paillier cryptosystem. We could adapt this encryption scheme of [37] to achieve fault tolerance, however the fault tolerance in terms of verifiability should allow the Aggregator to decrypt the result and compute the proof. As the tags computation employ a discrete log based group with randomness in the exponent, which resembles BLS signatures, any time a user does not encrypt and tag its data, then the key dealer should be aware of that user and distribute to the Data Analyzer new verification key. That involves the online existence of the key dealer, which would render the protocol insecure in terms of a key dealer compromisation. Another approach would be the adaptation of the aforementioned techniques with increased communication cost or error in the final result.

9 Comparison

9.1 Security

We present a detailed comparison with respect to the security model and the collusion resistant property of existing protocols in table 4. Protocols which assure *Aggregator obliviousness* (AO) protect individual privacy from semi-honest Aggregators. Interestingly, a recent published paper [35], necessitates the appropriate and rigorous security analysis that should be conducted for secure aggregation protocols. As already mentioned in [23] there are two flaws in [35]. By exploiting the underlying mathematical structure of the encryption algorithms a passive adversary can fully recover the plaintext values from the ciphertext of a user. Moreover collusions, which are allowed as stated in the trust model of the paper, permit users to annihilate the randomness used to evaluate multiplications over plaintexts. Apart from this flawed protocol, to the best of our knowledge all the existing protocols guarantee AO in case of collusions, simply because each user does not share the encryption key with any other party in the protocol, thus the Aggregator cannot distinguish individual ciphertexts. Verifiability allows a party in the protocol to verify the correctness of the results performed by a malicious Aggregator. The protocol in [38] achieves public verifiability with the assumption of trustworthy users. As we showed in section 2.2, this protocol is insecure with respect to unforgeability as long as a malicious user colludes with a malicious Aggregator. Verifiability is also achieved in Barthe *et al.* [6] but in a different context. The authors presented a tool-assisted verifiable computations framework for program code verification for differential private computations. Thus, the notion of collusions cannot be used in program verification code for comparison with our work. The rest of the protocols do not tackle the verifiability property with malicious Aggregators, apart from [20], which addresses integrity only protection on data and not verifiability in case of a malicious Aggregator. Our solution is not fault tolerant, but the core goal of our scheme is to enahnce the security model with malicious Aggregators and users, who will try to collude with the Aggregator.

9.2 Costs

We performed a thorough comparison in terms of theoretical computation and communication cost with the most relevant work of [45] and [38] (cf. table 5). The scheme in [38], assures aggregate verifiability without collusion resistance. The extra cost comes first at the user side: for the computation of the metatag sent to the converter and the proof computation during the zero knowledge subprotocol with the Converter. Converter does not exist in [38]. During Aggregation our scheme is charged with extra $(n-1)$ multiplications in $\mathbb{G}_T$ for the collusion resistance verifiability property. In contrast, for the verifiability property the proof computation in [38] takes $(n-1)$ multiplications in $\mathbb{G}_1$. Interestingly, the data analyzer in our scheme is engaged with one less pairing evaluations (two instead of three) compared with [38], and only two more exponentiations in $\mathbb{G}_T$. We also compare our scheme with the baseline [45], which assures only aggregate obliviousness and not verifiability. In all schemes the cost for decryption comes at the cost of $\sqrt{2^{l-1}}$. The cost comes from the square root Pollard kangaroo algorithm of finding discrete logarithms.

Protocol	Obliviousness	Verifiability	CR	FT
Shi *et al.* [45]	✓	✗	✓	✓
Joye *et al.* [34]	✓	✗	✓	✓
Erkin *et al.* [25]	✓	✗	✓	✓
Li *et al.* [39]	✓	✗	✓	✓
Jawurek *et al.* [33]	✓	✓	✗	✓
Kursawe *et al.* [36]	✓	✗	✓	✓
Barthe *et al.* [6]	✓	✓	-	✓
Leontiadis *et al.* [37]	✓	✗	✓	✓
Leontiadis *et al.* [38]	✓	✓	✗	✓
Günther *et al.* [31]	✓	✗	✗	✓
Melis *et al.* [42]	✓	✗	✓	✗
Bakondi *et al.* [4]	✓	✓	✗	✓
Bao *et al.* [5]	✓	✗	✓	✗
Chen *et al.* [21]	✓	✗	✓	✓
Patsakis *et al.* [43]	✓	✗	✓	✗
Won *et al.* [46]	✓	✗	✓	✗
Chen *et al.* [20]	✓	✗	✗	✓
This work	✓	✓	✓	✗

Table 4: Comparison of existing protocols. CR denotes whether a protocol is collusion resistant or not. FT is for fault tolerance

	This work		[38]		[45]	
Participant	Comp	Comm	Comp	Comm	Comp	Comm
User	$6E^{\mathbb{G}_1} + 1E^{\mathbb{G}_2} + 1E^{\mathbb{G}_T} + 2H + 1I^{\mathbb{Z}_p} + 3M^{\mathbb{G}_1} + 1P$	$6l + l_T$	$4E^{\mathbb{G}_1} + 1H + 2M^{\mathbb{G}_1}$	$2l$	$2E^{\mathbb{G}_1} + 1H + 1M^{\mathbb{G}_1}$	l
Converter	$4E^{\mathbb{G}_1} + 1E^{\mathbb{G}_T} + 1H + 2M + 1P$	$l + l_T$	-	-	-	-
Aggregator	$\sqrt{2^{l-1}} + (n-1)M^{\mathbb{G}_1} + (n-1)M^{\mathbb{G}_T}$	$l + l_T$	$\sqrt{2^{l-1}} + (2n-2)M^{\mathbb{G}_1}$	l	$\sqrt{2^{l-1}} + (n-1)M^{\mathbb{G}_1}$	-
Data Analyzer	$2E^{\mathbb{G}_T} + 1H^{\mathbb{G}_1} + 2M^{\mathbb{Z}_p} + 1M^{\mathbb{G}_T} + 2P$	-	$1E^{\mathbb{G}_1} + 1H + 1M^{\mathbb{G}_T} + 3P$	-	-	-

Table 5: Comparison of computation and communication cost with [38] and [45]. l denotes the bit-size of the prime number p and l_T the bit-size of elements in $\mathbb{G}_T$. E: exponentiation, I: inverse, H: hash, M: multiplication, P: pairing.

10 Conclusion

We addressed the problem of collusion resistant aggregation. Under this scenario users can collude with a malicious Aggregator, without the latter being able to forge other users' data. For our solution we initiate the study of *convertible tag*. Users first compute an authentication tag over their personal data and they forward this information along with some auxiliary data, which comprises a blinded version of their key, to an untrusted *Converter*. Finally the *Converter* transforms the tags, in order to allow an Aggregator to compute a proof of correct computations over user's data. We augment the current privacy definitions of Aggregate unforgeability with collusions between a user, the Aggregator and the *Converter*. Our protocol is provably secure and achieves constant time verification in the symmetric setting. Moreover, benchmark results justify its practicality.

References

1. J. A. Akinyele, C. Garman, I. Miers, M. W. Pagano, M. Rushanan, M. Green, and A. D. Rubin. Charm: a framework for rapidly prototyping cryptosystems. *Journal of Cryptographic Engineering*, 3(2):111–128, 2013.
2. G. Ateniese, J. Camenisch, M. Joye, and G. Tsudik. A practical and provably secure coalition-resistant group signature scheme. In *Advances in Cryptology – CRYPTO 2000*, pages 255–270. Springer Berlin Heidelberg, 2000.
3. G. Ateniese and S. Hohenberger. Proxy re-signatures: New definitions, algorithms, and applications. In *Proceedings of the 12th ACM Conference on Computer and Communications Security*, CCS '05, pages 310–319, New York, NY, USA, 2005. ACM.
4. B. G. Bakondi, A. Peter, M. Everts, P. Hartel, and W. Jonker. Publicly verifiable private aggregation of time-series data. In *Availability, Reliability and Security (ARES), 2015 10th International Conference on*, pages 50–59, Aug 2015.
5. H. Bao and R. Lu. A new differentially private data aggregation with fault tolerance for smart grid communications. *IEEE Internet of Things Journal*, 2(3):248–258, 2015.
6. G. Barthe, G. Danezis, B. Grégoire, C. Kunz, and S. Z. Béguelin. Verified computational differential privacy with applications to smart metering. In *2013 IEEE 26th Computer Security Foundations Symposium, New Orleans, LA, USA, June 26-28, 2013*, pages 287–301, 2013.
7. P. Bichsel, J. Camenisch, G. Neven, N. Smart, and B. Warinschi. Get shorty via group signatures without encryption. In J. Garay and R. De Prisco, editors, *Security and Cryptography for Networks*, volume 6280 of *Lecture Notes in Computer Science*,

pages 381–398. Springer Berlin Heidelberg, 2010.
8. M. Blaze, G. Bleumer, and M. Strauss. Divertible protocols and atomic proxy cryptography. In K. Nyberg, editor, *Advances in Cryptology – EUROCRYPT'98*, volume 1403 of *Lecture Notes in Computer Science*, pages 127–144. Springer Berlin Heidelberg, 1998.
9. A. Boldyreva, A. Palacio, and B. Warinschi. Secure proxy signature schemes for delegation of signing rights. *J. Cryptology*, 25(1):57–115, 2012.
10. D. Boneh, X. Boyen, and H. Shacham. Short group signatures. In M. Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 41–55. Springer Berlin Heidelberg, 2004.
11. S. A. Brands. An efficient off-line electronic cash system based on the representation problem. Technical report, Amsterdam, The Netherlands, The Netherlands, 1993.
12. J. Camenisch and A. Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *Advances in Cryptology – EUROCRYPT 2001*, pages 93–118. Springer Berlin Heidelberg, 2001.
13. J. Camenisch and A. Lysyanskaya. Signature schemes and anonymous credentials from bilinear maps. In M. Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 56–72. Springer Berlin Heidelberg, 2004.
14. J. Camenisch and M. Stadler. Efficient group signature schemes for large groups. In *Advances in Cryptology – CRYPTO'97*, pages 410–424. Springer Berlin Heidelberg, 1997.
15. J. Camenisch and M. Stadler. Efficient group signature schemes for large groups. In J. Kaliski, BurtonS., editor, *Advances in Cryptology – CRYPTO '97*, volume 1294 of *Lecture Notes in Computer Science*, pages 410–424. Springer Berlin Heidelberg, 1997.
16. T.-H. H. Chan, E. Shi, and D. Song. Privacy-preserving stream aggregation with fault tolerance. In *Financial Cryptography*, pages 200–214, 2012.
17. D. Chaum. Blind signatures for untraceable payments. In *CRYPTO*, pages 199–203, 1982.
18. D. Chaum. Blind signatures for untraceable payments. In D. Chaum, R. Rivest, and A. Sherman, editors, *Advances in Cryptology*, pages 199–203. Springer US, 1983.
19. D. Chaum and E. van Heyst. Group signatures. In D. Davies, editor, *Advances in Cryptology – EUROCRYPT 1991*, volume 547 of *Lecture Notes in Computer Science*, pages 257–265. Springer Berlin Heidelberg, 1991.
20. J. Chen, H. Ma, and D. Zhao. Private data aggregation with integrity assurance and fault tolerance for mobile crowd-sensing. *Wireless Networks*, pages 1–14, 2015.
21. L. Chen, R. Lu, and Z. Cao. Pdaft: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications. *Peer-to-peer networking and applications*, 8(6):1122–1132, 2015.
22. R. Cramer, R. Gennaro, and B. Schoenmakers. A secure and optimally efficient multi-authority election scheme. In *Proceedings of the 16th Annual International Conference on Theory and Application of Cryptographic Techniques*, EUROCRYPT'97, pages 103–118, Berlin, Heidelberg, 1997. Springer-Verlag.
23. A. Datta and M. Joye. Cryptanalysis of a privacy-preserving aggregation protocol. *IEEE Trans. Dependable Sec. Comput.*, 2016. http://joye.site88.net/papers/DJ__cryptanalysis.pdf.
24. D. Derler, C. Hanser, and D. Slamanig. Privacy-enhancing proxy signatures from non-interactive anonymous credentials. In V. Atluri and G. Pernul, editors, *Data and Applications Security and Privacy XXVIII*, volume 8566 of *Lecture Notes in Computer Science*, pages 49–65. Springer Berlin Heidelberg, 2014.
25. Z. Erkin and G. Tsudik. Private computation of spatial and temporal power consumption with smart meters. In *ACNS*, pages 561–577, 2012.
26. J. Fan, Q. Li, and G. Cao. Privacy-aware trustworthy data aggregation in mobile sensing. In *IEEE Conference on Communications and Network Security*, 2015.
27. A. Fiat and A. Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *Proceedings on Advances in cryptology—CRYPTO '86*, pages 186–194, London, UK, UK, 1987. Springer-Verlag.
28. G. Fuchsbauer and D. Pointcheval. Anonymous proxy signatures. In R. Ostrovsky, R. De Prisco, and I. Visconti, editors, *Security and Cryptography for Networks*, volume 5229 of *Lecture Notes in Computer Science*, pages 201–217. Springer Berlin Heidelberg, 2008.
29. S. D. Galbraith, F. Hess, and F. Vercauteren. Aspects of pairing inversion. *IEEE Trans. of Information Theory.*, 54:5719–5728, 2008.
30. K. Grining, M. Klonowski, and P. Syga. *Practical Fault-Tolerant Data Aggregation*, pages 386–404. Springer International Publishing, Cham, 2016.
31. F. Günther, M. Manulis, and A. Peter. Privacy-enhanced participatory sensing with collusion re-

sistance and data aggregation. In *Cryptology and Network Security - 13th International Conference, CANS 2014, Heraklion, Crete, Greece, October 22-24, 2014. Proceedings*, pages 321–336, 2014.
32. C. Hanser and D. Slamanig. Blank digital signatures. In *Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security*, ASIA CCS '13, pages 95–106, New York, NY, USA, 2013. ACM.
33. M. Jawurek and F. Kerschbaum. Fault-tolerant privacy-preserving statistics. In *Privacy Enhancing Technologies*, pages 221–238, 2012.
34. M. Joye and B. Libert. A scalable scheme for privacy-preserving aggregation of time-series data. In *Financial Cryptography*, 2013.
35. T. Jung, X. Li, and M. Wan. Collusion-tolerable privacy-preserving sum and product calculation without secure channel. *IEEE Trans. Dependable Sec. Comput.*, 12(1):45–57, 2015.
36. K. Kursawe, G. Danezis, and M. Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *PETS*, pages 175–191, 2011.
37. I. Leontiadis, K. Elkhiyaoui, and R. Molva. Private and dynamic time-series data aggregation with trust relaxation. In *Cryptology and Network Security - 13th International Conference, CANS 2014, Heraklion, Crete, Greece, October 22-24, 2014. Proceedings*, pages 305–320, 2014.
38. I. Leontiadis, K. Elkhiyaoui, M. Önen, and R. Molva. PUDA - privacy and unforgeability for data aggregation. In *Cryptology and Network Security - 14th International Conference, CANS 2015, Marrakesh, Morocco, December 10-12, 2015, Proceedings*, pages 3–18, 2015.
39. Q. Li and G. Cao. Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error. In *PETS*, pages 60–81, 2013.
40. B. Libert and D. Vergnaud. Multi-use unidirectional proxy re-signatures. In *Proceedings of the 15th ACM Conference on Computer and Communications Security*, CCS '08, pages 511–520, New York, NY, USA, 2008. ACM.
41. M. Mambo, K. Usuda, and E. Okamoto. Proxy signatures for delegating signing operation. In *Proceedings of the 3rd ACM Conference on Computer and Communications Security*, CCS '96, pages 48–57, New York, NY, USA, 1996. ACM.
42. L. Melis, G. Danezis, and E. D. Cristofaro. Efficient private statistics with succinct sketches. *CoRR*, abs/1508.06110, 2015.
43. C. Patsakis, P. Laird, M. Clear, M. Bouroche, and A. Solanas. Interoperable privacy-aware e-participation within smart cities. *Computer*, 48(1):52–58, 2015.
44. D. Pointcheval and J. Stern. Provably secure blind signature schemes. pages 252–265. Springer-Verlag, 1996.
45. E. Shi, T.-H. H. Chan, E. G. Rieffel, R. Chow, and D. Song. Privacy-preserving aggregation of time-series data. In *NDSS*, 2011.
46. J. Won, C. Y. Ma, D. K. Yau, and N. S. Rao. Proactive fault-tolerant aggregation protocol for privacy-assured smart metering. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 2804–2812. IEEE, 2014.