Collusion Resistant Aggregation from Convertible Tags

Iraklis Leontiadis  
University of Arizona  
leontiad@email.arizona.edu

Ming Li  
University of Arizona  
lim@email.arizona.edu

ABSTRACT
The progress in communication and hardware technology increases the computational capabilities of personal devices. Data is produced massively from ubiquitous devices that cannot be stored locally. Moreover, third party authorities in order to increase their value in the market with more knowledge, seek to collect individual data inputs, such that they can make a decision with more relevant information. 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 a honest-but-curious Aggregator who is trusted to follow the rules of the protocol or they model a malicious Aggregator with trustworthy users. That limits the security analysis to users who are trustworthy to not share any secret information with a malicious Aggregator. 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 under standard assumptions in the random oracle model.

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 or 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 patient according its health care data.

In the aforementioned use cases an untrusted Aggregator seeks to compute in cleartext a function f over users’ data. 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 [12, 18, 24, 28, 33] or on improving the security model with a malicious Aggregator who will try to convince a honest verifier that the result of computations comes from genuine data inputs. In [29] the authors by leveraging the encryption scheme of Shi et al. [33] they enrich the typical data collection and analysis protocol with a proof computed by a malicious Aggregator, which allows a verifier 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 [19] 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 security of the scheme. 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 \( f \). Formally, 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 [29] and Aggregator obliviousness [33]. That is, a malicious Aggregator cannot convince a honest verifier for the correctness of computations against a malicious Aggregator, without affecting the security of the scheme.

**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, in case of trustworthy users. 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 and in section 5 we identify some preliminaries. The protocol is presented in full details in section 6. Then, in section 7 we perform an analysis with respect to the security of the scheme, its costs and a comparison with existing protocols. Finally we conclude in section 8.

2. PROBLEM STATEMENT

For a secure aggregation protocol, we assume a set of \( n \) users \( U = \{ U_i \}_{i=1}^{n} \), 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 \( A \), whose main goal is to learn a function \( f \) in cleartext over users’ data. 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 a honest verifier 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 [29].

2.1 Syntax

- **Setup** \((1^\lambda) \rightarrow (pp, sk_A, \{ sk_i \}_{U_i \in U}, vk)\): It is a randomized algorithm run by a trusted dealer \( KD \), which on input of a security parameter \( \lambda \) outputs the public parameters \( pp \) that will be used by subsequent algorithms, the Aggregator \( A \)’s secret key \( sk_A \), the secret keys \( sk_i \) of users \( U_i \) and the public verification key \( vk \).

- **EncTag** \((t, sk_i, x_{i,t}) \rightarrow (c_{i,t}, \sigma_{i,t})\): It is a randomized algorithm which on inputs of time interval \( t \), secret key \( sk_i \) of user \( U_i \) and data \( x_{i,t} \) encyrpts \( x_{i,t} \) to get a ciphertext \( c_{i,t} \) and computes a tag \( \sigma_{i,t} \) that authenticates \( x_{i,t} \).

- **Aggregate** \((sk_A, \{ c_{i,t} \}_{U_i \in U}, \{ \sigma_{i,t} \}_{U_i \in U}) \rightarrow (\text{sum}, \sigma)\): It is a deterministic algorithm run by the Aggregator \( A \). It takes as inputs Aggregator \( A \)’s secret key \( sk_A \), ciphertexts \( \{ c_{i,t} \}_{U_i \in U} \) and authentication tags \( \{ \sigma_{i,t} \}_{U_i \in U} \), and outputs in cleartext the sum \( \text{sum}_t \) of the values \( \{ x_{i,t} \}_{U_i \in U} \). Moreover, it computes a proof \( \sigma \), assessing the correctness of \( \text{sum}_t \), using the authentication tags \( \{ \sigma_{i,t} \}_{U_i \in U} \).

- **Verify** \((vk, \text{sum}_t, \sigma) \rightarrow \{ 0, 1 \}\): It is a deterministic algorithm that is executed by the Data Analyzer \( DA \). It outputs 1 if Data Analyzer \( DA \) is convinced that the sum \( \text{sum}_t = \sum_{U_i \in U} \{ x_{i,t} \} \); and 0 otherwise, with the aid of the verification key \( vk \).

2.2 Security Model

We build upon the model as presented in [29] and we further assume that users are not trustworthy. 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 \( A \) cannot forge an aggregate tag with respect to the computation \( f \). That is, if users submit tags \( \sigma_{i,t} \) for their private data inputs \( x_{i,t} \), then \( A \) can only compute a valid aggregate tag \( \sigma_t \) for the sum computation over \( x_{i,t} \) and nothing else. We show how the scheme in [29] does not assure aggregate unforgeability (cf. Algorithms 3 and 4) in the presence of non-legitimate users, who collude with a malicious Aggregator \( A \). A malicious user \( U_m \) shares the secret information \( x \) with the Aggregator. The Aggregator \( A \) then can forge another user’s tag with \( x \) as follows: After obtaining \( \sigma_{i,t} = H(t)^{vk}(g_i^x)^{x_{i,t}} \) from a legitimate user \( U_i \) at time interval \( t \), \( A \) computes \( \sigma_{i,t} \cdot (g_i^v)^{\text{sum}_t} = H(t)^{vk}(g_i^x)^{x_{i,t} + \text{sum}_t} \), for a value \( v \) of its choice. Thus, \( A \) can produce a valid proof by aggregating all tags and the forged one, for a sum that comes from non-genuine data.

We do not consider re-ordering attacks as considered in the mix-and-match type of attacks [17], because order does not have an impact on the final verification information which is the sum over users’ 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 (cf. Algorithms 1 and 2).

3. RELATED WORK
Algorithm 1: Learning phase of the obliviousness game

\[
\begin{align*}
& (pp, sk_A, \{sk_i\}_{i \in U}, vk) \leftarrow \mathcal{O}_{\text{Setup}}(1^n) ; \\
& // A executes the following a polynomial number of times
& sk_i \leftarrow \mathcal{O}_{\text{Corrupt}}(uid_i) ; \\
& // A is allowed to call \mathcal{O}_{\text{EncTag}} for all users \mathcal{U}
& (c_{i,t}, \sigma_{i,t}) \leftarrow \mathcal{O}_{\text{EncTag}}(t, uid_i, x_{i,t}) ;
\end{align*}
\]

Algorithm 2: Challenge phase of the obliviousness game

\[
\begin{align*}
& A \rightarrow t^* , S^* ; \\
& A \rightarrow X_i^{\sigma_{i,t}} , X_t^{\sigma_t} ; \\
& (c_{i,t}^*, \sigma_{i,t}^*) _{i,t \in S^*} \leftarrow \mathcal{O}_{\text{AO}}(X_i^{\sigma_{i,t}}, X_t^{\sigma_t}) ; \\
& A \rightarrow b^* ;
\end{align*}
\]

Algorithm 3: Learning phase of the aggregate unforgeability game

\[
\begin{align*}
& pp, sk_A, \{sk_i\}_{i \in U}, VK \leftarrow \mathcal{O}_{\text{Setup}}(1^n) ; \\
& // A executes the following a polynomial number of times
& \mathcal{A} is allowed to call \mathcal{O}_{\text{EncTag}} for all users \mathcal{U}
& (c_{i,t}, \sigma_{i,t}) \leftarrow \mathcal{O}_{\text{EncTag}}(t, uid_i, x_{i,t}) ; \\
\end{align*}
\]

Algorithm 4: Challenge phase of the aggregate unforgeability game

\[
\begin{align*}
& (t^*, \sum_{c_{i,t} , \sigma_{i,t}}^*) \leftarrow \mathcal{A}
\end{align*}
\]

In practice the secret key of the original signer is split between the receiver and the proxy. Variations of proxy signatures as warrant-signatures [16, 21, 22] 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 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-signatures 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.

4. IDEA AND MODEL

4.1 Idea

The core idea (cf. figure 1) 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 C** and their ciphertexts to the malicious Aggregator \( A \). The protocol at this point assures unforgeability and obliviousness against the **Converter C**. Along with the metatags each user transmits to \( C \) some auxiliary information coupled with a blinded version of their secret tag key. \( C \) 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 \( A \). Users upon receiving their tags from \( C \)
validate its correctness. This is happening in order to ensure that in case of a collusion between a colluding user and the Converter \( C \), the latter cannot forward a forged tag, with the key that is used by \( C \) 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 \( A \), in order to forge an authentication tag for another user. Aggregator receives all tags and ciphertexts. \( A \) then decrypts and learns the result \( \sum_i = f = \sum_{i=1}^n x_{i,t} \) and computes a proof of correct computations \( \sigma_t \) based on the convertible tags.

Finally, \( A \) forwards to the data analyzer \( DA \) the result \( \sum_i \) and the proof \( \sigma_t \) that allows the data analyzer \( DA \) to verify whether or not \( C \) can be modeled as a malicious party, which deviates from the protocol’s rules. The assumption that \( C \) and \( A \) do not collude is possible to occur [26], in case of independent adversaries that have captured \( C \) and \( A \) with different motivations, or in case of increased costs. Finally collusions can be prevented in real world for the sake of positive business reputation.

As we extend the model for privacy preserving and unforgeable aggregation as presented in [29] 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

- **Setup\((1^\lambda)\)**: This is a probabilistic algorithm that on input the security parameter \( \lambda \) it outputs the public parameters \( pp \) and the secret key \( sk_A \) of the Aggregator.
- **UKeygen\((1^\lambda)/(KD, U)\)**: The key dealer \( KD \) runs this algorithm in order to distribute secret keys to each user for encryption. Moreover users choose uniformly at random their tag keys.
- **CKeygen\((1^\lambda)/(KD, U, C, DA)\)**: This key distribution algorithm runs between the users who blind their randomness from the **UKeygen\((1^\lambda)/(KD, U)\)** algorithm, send that to the Converter \( C \), and the latter distributes the secret authentication tag key to the data analyzer \( DA \).
- **EncTag\((pp, sk_i, x_{i,t})\)**: Each user using its secret encryption key encrypts its individual data and sends the ciphertext \( c_{i,t} \) to \( A \). Moreover using its secret tag key computes a metatag \( mtag_{i,t} \), and sends that to the Converter \( C \).
- **Convert\((pp, r, mtag_{i,t})\)**: Converts the tag \( mtag_{i,t} \) to \( C \) with the key \( r \) and the metatag \( mtag_{i,t} \) computes the final tag \( \sigma_{i,t} \) for user \( U_i \).
- **VTag\((pp, sk_i, \sigma_{i,t}, x_{i,t})\)**: Each user verifies the correctness of the final tag \( \sigma_{i,t} \). **Convertible tags** prevent \( C \) to forge a user’s tag using secret information from a colluding user.
- **Aggregate\((sk_A, \{c_{i,t}\}, \{\sigma_{i,t}\})\)**: Aggregator \( A \) upon collecting the ciphertexts \( \{c_{i,t}\} \) and the tags \( \{\sigma_{i,t}\} \) decrypts
with the secret key \( s_k_A \) and learns the result \( \sum_{t=1}^n x_{t,t} \). Moreover, it computes a proof of correct computation \( \sigma_t \) and finally and forwards to the data analyzer \( D_A \)

\[ \sum_{t=1}^n x_{t,t} \]

\[ \sigma_t \]

- **Verify** \((pp, vk, \sum_{t=1}^n, \sigma_t)\): The data analyzer \( D_A \) verifies the correctness of computation for the \( \sum_{t=1}^n \), using the proof \( \sigma_t \) 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 [29], our scheme fulfills its security guarantees under weakened assumptions. More specifically, collusions in between users and malicious parties are supported without jeopardizing privacy or security definitions. Users \( U = \{U_i\}_{i=1}^n \) in the scheme are unauthenticated and can act maliciously. Collusions can happen between a malicious user \( U_m \) and a colluding Aggregator \( A \) or a malicious Converter \( C \). Users share any secret information they know with the colluding members with the goal to forge other users’ tag. Collusions between \( C \) and \( A \) are also possible in case of trustworthy users (cf. table 1).

We first describe the oracles an adversary \( A \) has access to when collusions between \( U, C \) and \( A \) are possible:

- **\( O_{Coll:U.m} (uid = i \in U) \)**: On input a user identifier \( uid \) this oracle transmits to an adversary who impersonates a malicious Aggregator \( A \) the user’s secret information \( (ek_{uid}, tk_{uid}, uid, w) \) after running the \( UKeygen(1^\lambda) \) and \( CKeygen(1^\lambda) \) algorithms.

- **\( O_{Coll:C.m} (uid = i \in U) \)**: On input user identifier \( uid \) this oracle runs the \( UKeygen(1^\lambda) \) and \( CKeygen(1^\lambda) \) algorithms and forwards to a malicious Converter \( C \) the user’s secret information \( (ek_{uid}, tk_{uid}, uid, w) \).

- **\( O_{Coll:A.c} (uid = i \in U) \)**: In case of trustworthy users this oracle returns the secret key of \( C \) to an adversary \( A \).

#### 4.3.2 Collusion Resistant Aggregate Unforgeability

The security of the scheme is modeled under the collusion resistant aggregate unforgeability (\( CR - AU \)) security definition. An adversary \( A \) is able to obtain valid authentication tags for values of its choice by corrupting users. \( A \) also learns valid encryptions of its choice, and learns the final result over plaintext values \( \sum_{t=1}^n x_{t,t} \). In the end we claim that an aggregation scheme is secure if a malicious Aggregator \( A \) cannot forge an aggregate tag for a time interval \( t \) such that for the underlying plaintexts it holds that \( \sum_{t=1}^n x_{t,t} \neq \sum_{t=1}^n x_{t,t} \) for a set of users \( U_t \in S \) that did not collude with the Aggregator or the Converter. We follow the security syntax as in [29] and we differentiate between:

- **Type-I** forgeries in which \( A \) 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 \( A \) has received valid tags for the users but \( \sum_{t=1}^n x_{t,t} \neq \sum_{t=1}^n x_{t,t} \).

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 \( D_A \) for the correctness of computations given erroneous data inputs. An adversary during the CR-AU game has access to the following oracles:

- **\( O_{Setup} () \)**: This oracle when queried responds with the public parameters of the scheme \( pp \) and the secret key of the Aggregator \( s_k_A \).

- **\( O_{Coll:A.im} (uid = i \in U) \)**: On input a user identifier \( uid \), this oracle when is queried by a malicious Aggregator \( A \) replies with the secret key of a user \( sk_{uid} \).

- **\( O_{Coll:C.im} (uid = i \in U) \)**: Upon receiving a user identifier \( uid \) the \( O_{Coll:C.im} \) oracle responds to a malicious Converter \( C \) with the secret key of a user \( sk_{uid} \).

- **\( O_{Corr:A} () \)**: This oracle responds with the secret decryption key \( sk_A \) of the Aggregator.

- **\( O_{Corr:C} () \)**: This oracle responds with the secret key of the Converter \( C \).

- **\( O_{Corr:D.A} () \)**: This oracle responds with the secret verification key \( vk \) of the data analyzer \( D_A \).

- **\( O_{EncTag} (t, uid, x_{t,t}) \)**: This is an oracle that replies with the encryption of the value \( x_{t,t} \) using the secret key of the user \( i \) after calling the \( O_{Coll:A.im} (t, uid = i \in U) \) or \( O_{Coll:A.c} (t, uid = i \in U) \) oracle.

- **\( O_{Mtag} (mtag_{i,t}) \)**: The \( O_{Mtag} \) oracle on input a metatag \( mtag_{i,t} \) it converts it to the tag \( \sigma_{i,t} \) after corrupting Converter’s secret key with the \( O_{Corr:C} \) oracle.

- **\( O_{Aggregate} ((c_{i,t})^n_{i=1}) \)**: This oracle simulates the behavior of the Aggregator \( A \) and when invoked with inputs the ciphertexts \( (c_{i,t})^n_{i=1} \), it gives as a response the sum \( \sum_{t=1}^n x_{t,t} \), after calling the \( O_{Corr:A} \) oracle, in order to obtain the secret decryption key of the Aggregator \( sk_A \).

- **\( O_{Verify} (t, \sigma_t, sum_t) \)**: Upon receiving a tuple, containing a time interval \( t \), a proof \( \sigma_t \) and a result sum_t, the \( O_{Verify} \) oracle invokes the \( O_{Corr:D.A} \) oracle and replies with 1 \( \iff \sum_{t=1}^n x_{t,t} = \sum_{t=1}^n x_{t,t} \), or 0 otherwise.

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

In Game**\( ^{CR-AU-I} \) users are not trustworthy and collusions between a user and a malicious Aggregator \( A \) or a Converter \( C \) are allowed. During the learning phase of the game (cf. algorithm 5), \( A \) interacts with **\( O_{Setup} () \)**, **\( O_{Coll:A.im} (uid = i \in U) \)**, **\( O_{Coll:C.im} (uid = i \in U) \)**, **\( O_{EncTag} (t, uid, x_{t,t}) \)**, **\( O_{Mtag} (mtag_{i,t}) \)**, **\( O_{Verify} (t, \sigma_t, sum_t) \)** 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 **\( O_{Verify} (t, \sigma_t, sum_t) \)** \( A \) has access to the verification oracle. Note, that this oracle during the game makes sense since, our scheme operates in a symmetric setting, thus \( A \) cannot publicly verify. Finally, \( A \) outputs a forgery for a time interval \( t \). The forgery is successful if **\( Verify (pp, vk, \sum_{t=1}^n, \sigma_t) = 1 \)** for a time interval \( t \) in which \( A \) did not query the \( O_{EncTag} \) (Type-I forgery), or for \( t \) in which \( A \) called \( O_{Mtag} \) (Type-II forgery) and none of users \( U_t \in S \) collude with the Aggregator or the Converter.
Algorithm 5: Learning phase of the CR – AU – I game

\[(pp, sk_A) \leftarrow \mathcal{O}_{\text{Setup}}(1^\lambda); \]
// A executes the following a polynomial number of times
\[\mathcal{O}_{\text{CollA},i}\text{Enc}(uid = i \in U); \]
\[\mathcal{O}_{\text{CollI},i}\text{Enc}(uid = i \in U); \]
// A is allowed to call \(\mathcal{O}_{\text{EncTag}}\) for all users \(U\)
\[(c_{i,t}, \sigma_{i,t}) \leftarrow \mathcal{O}_{\text{EncTag}}(t, uid, x_{i,t}); \]
\[\mathcal{O}_{\text{Blind}}(\text{mtag}_{i,t}); \]
\[\mathcal{O}_{\text{Verify}}(t, \sigma_i, \sum_t); \]

Algorithm 6: Challenge phase of the CR – AU – I game

\[(t^*, \sum_{t^*}, \sigma_{t^*}) \leftarrow A \]

Definition 1. (CR – AU – I) An aggregation scheme is CR – AU – I secure if any probabilistic polynomial time adversary \(A\) has negligible probability \(\epsilon(\lambda)\) on the winning probabilities
\[\Pr[A_{\text{CR-AU-I}}(\lambda)] \leq \epsilon(\lambda). \]

In Game\(^{\text{CR-AU-I}}\) users are assumed as trustworthy and collusions between \(C\) and \(A\) can occur. During the security game though, in the learning phase (cf. algorithm 7) \(A\) does not have access to the \(\mathcal{O}_{\text{CollA},i}\text{Enc}(uid = i \in U)\) and \(\mathcal{O}_{\text{CollI},i}\text{Enc}(uid = i \in U)\) oracles during which users share their secret keys with \(C\) and \(A\). However, \(A\) has access to \(\mathcal{O}_{\text{CollA},i}\text{Enc}(uid = i \in U)\) oracle since Aggregator and Converter can collude. Similarly with Game\(^{\text{CR-AU-II}}\) \(A\) succeeds if it outputs during the challenge phase (cf. algorithm 8) either a Type-I or Type-II forgery.

Correspondingly for a scheme with trustworthy users we define:

Definition 2. (CR – AU – II) An aggregation scheme is CR – AU – II secure if any probabilistic polynomial time adversary \(A\) has negligible probability \(\epsilon(\lambda)\) on the winning probabilities
\[\Pr[A_{\text{CR-AU-II}}(\lambda)] \leq \epsilon(\lambda). \]

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 \(G_1, G_2, G_T\) be cyclic groups of large prime order \(p\) and \(g_1, g_2\) generators of \(G_1, G_2\) 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 \(g_1, g_2 \in G_1 \times G_2\) and \(a, b \in \mathbb{Z}_p\).

2. Computability: there exists an efficient algorithm that computes \(e(g_1^a, g_2^b)\) where \(g_1, g_2 \in G_1 \times G_2\) 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 \(e : (G_1 \times G_2) \rightarrow G_T\) be a bilinear pairing, \(g\) a generator of \(G_1\) and \(g_2\) a generator of \(G_2\) and \(p\) the order of \(G_1, G_2\) and \(G_T\). Given \(U = (g, g^a, g_2^b, g_2^c) \in G_1 \times V = (g_2, g_2^b, g_2^c) \in G_2\) for random \(a, b, c \in \mathbb{Z}_p\), we say that BCDH holds if the probabilities of a probabilistic polynomial time adversary \(A\) to compute \(W = e(g_1, g_2)^{ab}\) are negligible on input the security parameter \(\lambda\): \(\Pr\{W \leftarrow A(U, V)\} \approx 1\).

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

Let \(e : (G_1 \times G_2) \rightarrow G_T\) be a bilinear pairing, \(g\) a generator of \(G_1\) and \(g_2\) a generator of \(G_2\) and \(p\) the order of \(G_1, G_2\) and \(G_T\). We say that XDH holds if the probabilities of a probabilistic polynomial time adversary \(A\) to solve DDH and DL in \(G_1\) are negligible on input the security parameter \(\lambda\).
Algorithm 9: Learning phase of the Aggregator obliviousness game

\[(\text{pp, sk}_{\text{A}}, \text{vk}) \leftarrow O_{\text{Setup}}(1^\lambda);\]
\[O_{\text{Coll},i,t}(\text{uid} = i \in U);\]
\[O_{\text{Coll},c}(\text{uid} = i \in U);\]

// A executes the following a polynomial number of times
// A is allowed to call \(O_{\text{Enc},t}\) for all users \(U\)
\[(c_{i,t}, \sigma_{i,t}) \leftarrow O_{\text{Enc},t}(t, \text{uid}_i, x_{i,t});\]
\[O_{A}\{\text{mtag}_{i,t}\};\]
\[O_{A}(t, \sigma_t, \text{sum}_t);\]

Algorithm 10: Challenge phase of the Aggregator obliviousness game
\[A \rightarrow t^*, S^*;\]
\[A \rightarrow \text{A}^{0^*, \text{A}^1};\]
\[(c_{i,t}^*, \sigma_{i,t}^*)_{i,t \in U} \leftarrow O_{A}(\text{A}^{0^*, \text{A}^1});\]
\[A \rightarrow b^*;\]

6. PROTOCOL

In order to guarantee AO our protocol employs Shi et al. scheme [33]. For the sake of completeness we briefly describe their encryption scheme.

6.1 Shi-Chan-Rieffel-Chow-Song Scheme

- **Setup(1^\lambda)**: On input the security parameter \(\lambda\) this probabilistic algorithm outputs a cryptographic secure hash function \(H : \{0, 1\}^* \rightarrow G_1\), for a group \(G_1\) of large prime order \(p\). Through a secure channel the trusted key dealer \(KD\) distributes to each user a secret encryption key \(ek_i \in Z_p\), which is chosen uniformly at random. \(KD\) also forwards to the \(A\) the secret decryption key \(sk_{\text{A}} = \sum_{i=1}^n ek_i\).

- **Encrypt(ek_{i,t}, x_{i,t})**: To encrypt data value \(x_{i,t}\) at time interval \(t\) with secret key \(ek_i\), user \(U\) computes the ciphertext \(c_{i,t} = H(t)^{sk_{i,t}} g^{x_{i,t}} \in G_1\).

- **Aggregate\{\{c_{i,t}\}_{i,t \in U}, \{\sigma_{i,t}\}_{i,t \in U}, sk_{\text{A}}\}**: Upon receiving all the ciphertexts \(\{c_{i,t}\}_{i,t=1}^n\), the Aggregator computes: \(V_\text{t} = \left(\prod_{i=1}^n c_{i,t}\right) H(t)^{-sk_{\text{A}}} = H(t)^{-\sum_{i=1}^n sk_i g^{\sum_{i=1}^n x_{i,t}}} H(t)^{-\sum_{i=1}^n ek_i} = g^{\sum_{i=1}^n x_{i,t}} \in G_1\). Then \(A\) learns the sum \(\text{sum}_t = \sum_{i=1}^n x_{i,t} \in Z_p\) by computing the discrete logarithm of \(V_\text{t}\) on the base \(g\). 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 \(A\).

- **Setup_1(1^\lambda)**: On input the security parameter \(\lambda\) this probabilistic algorithm defines a cryptographic secure hash function \(H : \{0, 1\}^* \rightarrow G_1\), a bilinear pairing \(e : G_1 \times G_2 \rightarrow G_T\) of prime order \(p\) with generator \(g\). Finally it outputs the public parameters \(pp = (H, e, g, g_2)\). It also calls the **Setup(1^\lambda)** algorithm of the Shi et al. scheme and outputs the secret key of the Aggregator \(sk_{\text{A}}\).

- **UKeygen(1^\lambda) (KD, U)**: Each user independently chooses uniformly random tag keys \(tk_i\) and \(r_i\). Through a secure channel each \(U_i\) forwards \(r_i\) to the key dealer \(KD\), who computes \(\sum_{i=1}^n r_i\).

- **CKeygen(1^\lambda) (KD, U)**: The key dealer chooses uniformly at random a key \(r \in Z_p\) and a random generator \(w \in G_2\). It distributes through a secure channel \(r\) to the Converter \(C\). It also sends to the \(DA\) the secret verification key \(vk = (w, r, \sum_{i=1}^n r_i)\). Moreover it forwards \(w\) to each user. Then the key dealer \(KD\) goes off-line.

- **EncTag(pp, sk_{i,t}, x_{i,t})**: This deterministic algorithm takes as input the secret key of each user \(sk_i = (r_i, w, tk_i, ek_i)\) and the private values \(x_{i,t}\) and outputs the metatag:

\[\text{mtag}_{i,t} = (\text{mtag}_{i,t}^1, \text{mtag}_{i,t}^2) = (\{H(t)^{g^{x_{i,t}}} w\}^r, w^{1})\]

Moreover, users encrypt their data with the encryption key \(ek_i\), with the encryption scheme of Shi et al. [33] as already presented in 6.1. Finally, \(U_i\) forwards \(c_{i,t}\) to the Aggregator \(A\) and the metatag \(\text{mtag}_{i,t}\) to the Converter.

- **Converter(pp, r, 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 \(\text{mtag}_{i,t}\) and outputs the tag \(\sigma_{i,t}\) as follows:

\[\sigma_{i,t} = e(\text{mtag}_{i,t}^1, \text{mtag}_{i,t}^2)^r = e(\{H(t)^{g^{x_{i,t}}} w\}^r, w^{1})^r = e(H(t)^{g^{x_{i,t}}} w)^r e(g^{x_{i,t}}, w)^r = e(H(t)^{g^{x_{i,t}}} w)^r e(g^{x_{i,t}}, w)^r\]

- **Aggregate\{sk_{\text{A}}, \{c_{i,t}\}, \{\sigma_{i,t}\}\}**: The Aggregator \(A\) after collecting all the ciphertexts \(c_{i,t}\) for the users \(U\) decrypts with the secret key \(sk_{\text{A}}\) and learns the sum \(\sum_{i=1}^n x_{i,t}\). For the decryption algorithm \(A\) uses the decryption algorithm as in Shi et al. scheme [33]. Moreover, \(A\) computes a proof of correct computation by aggregating the tags \(\sigma_{i,t}\) as follows:

\[\sigma_t = \prod_{i=1}^n \sigma_{i,t} = \prod_{i=1}^n e(\text{mtag}_{i,t}^1, \text{mtag}_{i,t}^2)^r = \prod_{i=1}^n e(H(t)^{g^{x_{i,t}}} w)^r e(g^{x_{i,t}}, w)^r = \prod_{i=1}^n e(H(t)^{g^{x_{i,t}}} w)^r \prod_{i=1}^n e(g^{x_{i,t}}, w)^r = e(H(t)^{g^{x_{i,t}}} w)^r e(g(w), w)^r e(\sum_{i=1}^n x_{i,t})\]

Finally \(A\) returns to the honest veriﬁer the result \(\text{sum}_t = \sum_{i=1}^n x_{i,t}\) and the proof \(\sigma_t = e(H(t), vk_1) \sum_{i=1}^n e(g, vk_1)^{2\text{sum}_t} = \sigma_t\)

- **Verify_1(pp, vk, sum_t, \sigma_t)**: The data analyzer \(DA\), who acts as honest veriﬁer veriﬁes the correctness of the sum computation by employing its veriﬁcation key \(vk = (vk_1 = w, vk_2 = r, vk_3 = \sum_{i=1}^n r_i)\). \(DA\) veriﬁes by checking if the following equation holds:

\[e(H(t), vki)^{2vk_3} e(g, vki)^{2\text{sum}_t} = \sigma_t\]
Thanks to the bilinearity of the pairings the correctness of the verification procedure is assured. Indeed:
\[
e(H(t), v)_{\sum_{i=1}^{n} e(g(w), v)}^{v_{k2\sum_{i=1}^{n}}} = e(H(t), w)^{\sum_{i=1}^{n} e(g(w), v)(x_{i,t})} = \sigma_t
\]

6.3 Collision resistant aggregation II (CRA-II)

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

**Attack on CRA-I scheme** A colluding user \(U_t\) shares with the Converter his secret tag key \(t_k\) and the shared common key between all users. \(C\) can forge a valid tag \(t_{ag,i}t\) for a trustworthy user \(U_t\) as follows: \(t_{ag,i,t} = e(g(t_i,t), w) = e(H(t), w)^{\sum_{i=1}^{n} e(g(w), v)(x_{i,t})}, \) which is a valid for the value \(x_{i,t} + x_{i,t}'\).

The core idea to mitigate these type of attacks is to enforce the Converter \(C\) to re-randomize the metatag \(t_{ag,i,t}\) = \(H(t)^{g(x_{i,t})}r\), with the randomness \(r\), such that \(C\) replies to \(U_t\) with the final tag \(\sigma_t = e(H(t), w)^{\sum_{i=1}^{n} e(g(w), v)(x_{i,t})} \) along with the randomized metatag \(t_{ag,i,t} = [H(t)^{g(x_{i,t})}]^{r}\). 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 collisions between a malicious user and a malicious \(C\), the latter can forge the final tag, but the user can detect it, thanks to the unforgeability of the metatag \(t_{ag,i,t}\). We describe the entire protocol for collusion resistant aggregation against \(A\) and \(C\):

- **Setup_{II}(1^\lambda)**: This algorithm calls the Setup_{I}(1^\lambda) algorithm and outputs the public parameters pp = \((H, e, g, g_2)\) and the secret key of the Aggregator sk_A

- **UKeygen_{II}(1^\lambda)(KD, \mathbb{U})**: UKeygen_{II}(1^\lambda) invokes the UKeygen_{I}(1^\lambda) algorithm during which each user independently chooses uniformly random tag keys \(t_k\) and \(r\). Moreover users transmit \(r_i\) through a secure channel to the key dealer who computes \(\sum_{i=1}^{n} r_i\).

- **CKeygen_{II}(1^\lambda)(KD, \mathbb{U}, C, DA)**: This algorithm calls the CKeygen_{I}(1^\lambda)(KD, \mathbb{U}, C, DA), in which the key dealer outputs the secret verification key \(vk = (w, r, \sum_{i=1}^{n} r_i)\), chooses uniformly at random a key \(r \in \mathbb{Z}_p\) and a random generator \(w \in G_2\). It distributes through a secure channel \(r\) to the Converter \(C\). It also sends \(w, r\) to each user, and forwards the result \(\sum_{i=1}^{n} x_{i,t}\) to the data analyzer \(D_{A}\).

- **EncTag_{II}(pp, sk_i, x_{i,t})**: EncTag_{II}(pp, sk_i, x_{i,t}) calls EncTag_{I}(pp, sk_i, x_{i,t}) and operates similarly. It outputs for each user \(U_t\) the ciphertext \(c_{i,t}\) and the metatag:

\[
tag_{i,t} = (mtag_{i,t}^1, mtag_{i,t}^2) = ([H(t)^{g(x_{i,t})}]^{r}, w_{\overline{r}_{i,t}})
\]

which is forwarded to the Converter \(C\).

- **Convert_{II}(pp, r, mtag_{i,t})**: Upon receiving the metatag \(mtag_{i,t} = (mtag_{i,t}^1, mtag_{i,t}^2) = ([H(t)^{g(x_{i,t})}]^{r}, w_{\overline{r}_{i,t}}), C\)

uses its secret key \(r\) to compute the final tag as follows:
\[
\sigma_{i,t}^1 = e(mtag_{i,t}^1, mtag_{i,t}^2)^r = e(H(t)^{g(x_{i,t})} r, w_{\overline{r}_{i,t}})^r = e(H(t)^{r}, w_{\overline{r}_{i,t}})^r = e(H(t)^{r}, w)^r e(g^{x_{i,t}}, w)^r
\]
\[
\sigma_{i,t}^2 = (mtag_{i,t}^1)^r = [H(t)^{g(x_{i,t})}]^{r}
\]

Finally \(C\) sends to \(U_t\) the final tag \(\sigma_{i,t} = (\sigma_{i,t}^1, \sigma_{i,t}^2)\).

- **Verify_{II}(pp, vk, sum_{i}, \sigma_t)**: Each user verifies the correctness of the final tag as follows:
\[
e(\sigma_{i,t}^2, w)^\frac{1}{r} = \sigma_{i,t}^2
\]

The correctness of the equation holds since:
\[
e(\sigma_{i,t}^2, w)^\frac{1}{r} = e(H(t)^{g(x_{i,t})} r, w_{\overline{r}_{i,t}})^r = e(H(t)^{r}, w)^r e(g^{x_{i,t}}, w)^r
\]
\[
\sigma_{i,t}^2 = (mtag_{i,t}^1)^r = [H(t)^{g(x_{i,t})}]^{r}
\]

At this point if the equation is not true the user \(U_t\) halts the execution of the protocol and it infers that \(C\) forged the tag \(\sigma_{i,t}\). Otherwise it continues by sending the final tag \(\sigma_{i,t} = (\sigma_{i,t}^1, \sigma_{i,t}^2)\) to the Aggregator \(A\).

- **Aggregate_{II}(sk_A, \{c_{i,t}\}, \{\sigma_{i,t}\})**: This algorithm calls Aggregate_{I}(sk_A, \{c_{i,t}\}, \{\sigma_{i,t}\}), which consecutively decrypts with the secret key \(sk_A\) and \(A\) learns \(\sum_{i=1}^{n} x_{i,t}\). Moreover, it computes a proof of correct computation \(\sigma_t\) and finally and forwards the result \(\sum_{i=1}^{n} x_{i,t}\) and the proof \(\sigma_t = \prod_{i=1}^{n} \sigma_{i,t} = e(H(t), w)^{\sum_{i=1}^{n} e(g(w), v)(x_{i,t})} \) to the data analyzer \(D_{A}\).

- **Verify_{II}(pp, vk, sum_{i}, \sigma_t)**: The Verify_{II}(pp, vk, sum_{i}, \sigma_t) algorithm invokes Verify_{I}(pp, vk, sum_{i}, \sigma_t) and verifies the correctness of the sum computation by checking:
\[
e(H(t), vk_{1})^{v_{k2\sum_{i=1}^{n}}} e(g(vk_{1}), vk_{1})^{v_{k2\sum_{i=1}^{n}}} = \sigma_t
\]

7. 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. [33] in case of corrupted users, thanks to the trusted key dealer that distributes individual secret keys to each user. As such, we assume a trusted key distribution phase before the key dealer \(KD\) goes off-line.

7.1 Aggregator Obliviousness

**Theorem 1.** The CRA-I and CRA-II schemes provide Aggregator Obliviousness under the DDH assumption in \(G_1\) in the random oracle mode.

**Proof.** We assume an adversary \(A\) 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 \(B\) invokes \(A\) as a subroutine in order to break the Aggregator obliviousness definition as defined in the scheme of
Shi et al. [33]. We will refer to this scheme as private streaming aggregation (PSA). Adversary $B$ has access to $O_{PSA}^{\text{Setup}}, O_{PSA}^{\text{Corrupt}},$ \text{Corrupt}, and $O_{PSA}^{\text{Encrypt}},$ and the challenger, when she tries to break $AO$ in PSA. The $O_{PSA}^{\text{Setup}}$ oracle gives the public parameters and the secret keys to the users and the Aggregator. The $O_{PSA}^{\text{Encrypt}}$ oracle on input a user id $i$ returns the secret encryption key $sk_i$ of a corrupted user. The $O_{PSA}^{\text{Encrypt}}$ oracle on input a data input $x_{i,t}$ returns the encryption $c_{i,t}$ under the encryption algorithm of [33].

The $O_{PSA}^{\text{Corrupt}}$ oracle during the challenge phase with $B$ flips a random coin $b \leftarrow \{0, 1\}$ and responds with the encryption of the time series $X^b_i = \{x_{i,t}\}$. Algorithm $B$ simulates as a challenger the oracles $A$ has access to during the Learning phase as follows:

- $O_{PSA}^{\text{Setup}}(\lambda)$: Whenever $A$ calls the $O_{PSA}^{\text{Setup}}(\lambda)$ oracle, $B$ calls the $O_{PSA}^{\text{Setup}}$ oracle, which responds to $B$ with a hash function $H : \{0, 1\}^* \rightarrow \mathbb{G}_1$, a generator $g$ of the group $\mathbb{G}_1$ of safe prime order $p$, and the Aggregator’s secret key $sk_A = \sum_{i \in \mathbb{Z}_p} ek_i$. Moreover, $B$ chooses the parameters of a bilinear pairing $e = (e, g_1, g_2, \mathbb{G}_1, \mathbb{G}_2)$. Uniformly at random it selects secret keys $r_i, \{r_i\} \overset{\text{r}}{\leftarrow} \mathbb{Z}_p, w \overset{\text{r}}{\leftarrow} \mathbb{G}_2$. Finally $B$ replies to $A$ with $H, g, b, sk_A$.

- $O_{PSA}^{\text{Corrupt}}(\text{uid})$: When $A$ invokes this oracle then $B$ calls the $O_{PSA}^{\text{Corrupt}}$ oracle and transmits to $A$ the secret encryption key $ek_i$ of a corrupted user $U_i \in S$ and its secret tag key $r_i, w_i$. \text{Corrupt}\(\text{uid}\(\text{uid}\)}(\text{uid}\(\text{uid}\))$. When $A$ invokes this oracle then $B$ calls the $O_{PSA}^{\text{Corrupt}}$ oracle and transmits to $A$ the secret encryption key $ek_i$ of a corrupted user $U_i \in S$ and its secret tag key $r_i, w_i$.

- $O_{PSA}^{\text{Encrypt}}(\text{uid})$: The collusion between the Converter and $A$ are simulated by the $O_{PSA}^{\text{Encrypt}}(\text{uid})$ oracle. $B$ replies on these calls with the secret $r_i, w_i$.

- $O_{PSA}^{\text{Encrypt}}(\text{uid}, x_{i,t})$: Upon call on the $O_{PSA}^{\text{Encrypt}}(\text{uid}, x_{i,t})$ oracle, $B$ invokes the $O_{PSA}^{\text{Encrypt}}$ with input $(t, \text{uid}, x_{i,t})$, which in turn replies to $B$ with the encryption $c_{i,t} = H(t)^{ek_i} g_1^{x_{i,t}}$ of $x_{i,t}$. $B$ also computes $mtag_{i,t} = c_{i,t}^r, w^t_i = (H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i$. Notice that $mtag_{i,t}$ is indistinguishable from the random one if we interchange the randomness and set $r_i = ek_i$ and $t_i = r_i, w_i$ for uniformly random keys $ek_i, r_i$. $B$ finally replies to $A$ with $(c_{i,t}, mtag_{i,t})$.

- $O_{PSA}^{\text{Tag}}(mtag_{i,t})$: $A$ calls this oracle in order to learn the final tag of each user $c_{i,t}$. $B$ computes the final tag as $\sigma_i = e(mtag_{i,t}^1, mtag_{i,t}^2) = e(H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i)^t = e(H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i)^t$. Under the verification key $vk = (w, r, sk_A)$ the aggregation of the tags $\prod_{i \in \mathbb{Z}_p} \sigma_i$ can be correctly verified, using the call $O_{PSA}^{\text{Verify}}(t, \sigma_i, \text{sum}_i)$ oracle.

- $O_{PSA}^{\text{Verify}}(t, \sigma_i, \text{sum}_i)$: $A$ can query this oracle to learn the result of verification. We assume a honest verifier and this oracle makes sense, since we are in a symmetric verifications setting with a secret key. $B$ returns the result of the verification since it knows the secret verification key $vk = (vk_1, vk_2, vk_3) = (w, r, sk_A)$:

$$e(H(t), vk_1)^{wk_2} e(g, vk_1)^{vk_3} = \sigma_t$$

When the learning phase is over, then $A$ during the Challenge phase, chooses a set of users $i \in \mathcal{S}$, that have not been corrupted during the Learning phase and chooses two time series $X^1_i = \{u_{i,t}^1, t^1, e_{i,t}^1, c^1_{i,t}\}$ and $X^2_i = \{u_{i,t}^2, t^2, e_{i,t}^2, c^2_{i,t}\}$ such that $\sum_{t^2} x^2_{i,t} = \sum_{t^1} x^1_{i,t}$ for a time interval $t^*$ in which $A$ did not query neither the $O_{\text{EncTag}}$ for the $O_{\text{Mtag}}$ oracle and sends them to $O_{AO}(\lambda^0, \lambda^1)$ oracle.

To simulate $O_{AO}(\lambda^0, \lambda^1)$ $B$ queries the $O_{PSA}^{\text{Encrypt}}$ oracle with input $X^0_i, X^1_i$, which in turn flips a random coin $b \leftarrow \{0, 1\}$ and responds to $B$ with the ciphertexts $(c^b_{i,t}, \text{uid})_{i \in \mathcal{E}^b}$. $B$ also computes the final tags:

$$\sigma^b_{i,t} = e(c^b_{i,t}, H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i)^t (1)$$

$$= e(H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i)^t (2)$$

Finally $B$ forwards $(c^b_{i,t}, \sigma^b_{i,t})$ to $A$. The tag $\sigma^b_{i,t}$ simulates perfectly the final tag of a user and the aggregation of the tags for the computation of the final proof $\sigma_t$ correctly verifies the sum under the secret verification key $vk = (vk_1, vk_2, vk_3) = (w, r, sk_A)$:

$$\prod_{i=1}^{n} \prod_{t=1}^{n} e(H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i)^t = e(H(t)^{ek_i} g_1^{x_{i,t}})^t, w^t_i)^t (3)$$

$$= \sigma_t = e(H(t), vk_1)^{vk_2} e(g, vk_1)^{vk_3} (4)$$

If $A$ has non-negligible advantage $\epsilon$ to correctly guess the bit $b^*$ for the bit $b$, then $B$ will break the AO game in the PSA scheme with non-negligible advantage $\epsilon$. 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 assures the intractability of DDH in $\mathbb{G}_1$.

\section{7.2 Aggregate unforgeability}

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

We will prove the following theorem in three steps. First we prove the security of a base scheme (BaseLine) without any collisions in between a user and any other party. To model this scheme, an adversary $A$ plays the game as described in algorithms 5 and 6 without access to the corruption oracles $O_{PSA}^{\text{Corrupt}}, O_{PSA}^{\text{EncTag}}$ and $O_{PSA}^{\text{Mtag}}$. For the sake of clarity we call the security definition of aggregate unforgeability in the BaseLine scheme as BAU and the corresponding game $\text{Game}_{BAU}$. Then we show that a \textit{Type-I} forgery in the CRA-I can be transformed to a \textit{Type-I} forgery in the BaseLine scheme and finally that a \textit{Type-I} forgery in the CRA-II scheme can be transformed to a \textit{Type-I} forgery in the BaseLine scheme, as well.

\textbf{Lemma 1}. The baseline scheme guarantees aggregate unforgeability for \textit{Type-I} forgeries under the BCDH assumption in the random oracle model.

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

- $B$ first chooses uniformly at random secret keys $w, t, \{r_i, ek_i, tk_i\}_{i=1}^{n}$.

\textbf{Learning phase}:

- $O_{\text{Setup}}$: Whenever $A$ calls this oracle, $B$ returns the public parameters $pp = (H, e, g, g_2)$ 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 \( sk_A = \sum_{i=1}^{n} ek_i. \) \( B \) also sets as the secret verification key \( \nu = (g_2^\nu, t, g^\nu \sum_{i=1}^{n} r_i) \) and does not share this information.

- \( A \) can query the random oracle \( H \) for a time interval \( t. \) In order to respond to the queries \( B \) constructs a list \( R_i(t : v_t, \text{coin}(t), H(t)) \) and responds to \( A \) as follows:
  - If \( H \) has been queried before at the time interval \( t, \) \( B \) fetches the tuple \( R_i(t) \) and replies to \( A \) with \( H(t). \)
  - If \( t \) is fresh then \( B \) selects uniformly at random \( \phi_t \in \mathbb{Z}_p \) and flips a random \( \text{coin}(t). \) With probability \( p \) \( \text{coin}(t) = 0 \) and \( B \) appends to \( R_i(t) = g_2^{\phi_t}. \) Otherwise with probability \( 1 - p \) when \( \text{coin}(t) = 1 \) then \( B \) sets \( R_i(t) = g^\nu \phi_t. \) Finally \( B \) sends \( R_i(t) \) to \( A. \)
  - Whenever \( A \) calls the \( O_{\text{EncTag}}(t, \text{uid}, x_{i,t}) \) oracle, \( B \) constructs a tuple \( \text{ET}(t, \text{uid}, x_{i,t}, \sigma_{i,t}). \)

We have three cases:

1. If at time interval \( t, \) \( O_{\text{EncTag}}(t, \text{uid}, x_{i,t}) \) has not been queried before, then \( B \) calls the simulated random oracle for time interval \( t \) and gets the response \( H(t). \) If \( \text{coin}(t) = 1 \) then \( B \) halts the simulation. Otherwise it computes the ciphertext with the secret encryption key \( ek_i, \) as \( c_{i,t} = H(t)^{ek_i} g^{\nu \phi_t}. \) Finally \( B \) compiles the metatag \( \text{mtag}_{i,t} = [H(t)^{i} g^{\nu \phi_t}, w_i^2, t]^2 \) forwards to \( c_{i,t}, \text{mtag}_{i,t} \) to \( A. \) It also updates ET list with the tuple: \( \langle t, \text{uid}, x_{i,t}, \sigma_{i,t} \rangle \) and sets \( \Sigma_t = \Sigma_t + x_{i,t}. \)

2. If there exists \( \text{uid} \) in the list ET for time interval \( t, \) then \( B \) fetches this tuple and forwards \( c_{i,t}, \sigma_{i,t} \) to \( A. \)

3. Else \( B \) fetches the corresponding tuple from the \( R_i \) list. If \( \text{coin}(t) = 1 \) then \( B \) halts the simulation. Otherwise it computes the ciphertext with the secret encryption key \( ek_i, \) as \( c_{i,t} = H(t)^{ek_i} g^{\nu \phi_t}. \) Finally \( B \) compiles the metatag \( \text{mtag}_{i,t} = [H(t)^{i} g^{\nu \phi_t}, w_i^2, t]^2 \) forwards to \( c_{i,t}, \text{mtag}_{i,t} \) to \( A. \) It also updates ET list with the tuple: \( \langle t, \text{uid}, x_{i,t}, \sigma_{i,t} \rangle \) and sets \( \Sigma_t = \Sigma_t + x_{i,t}. \)

- When \( A \) calls the \( O_{\text{Tag}}(\text{mtag}_{i,t}) \) oracle, \( B \) calls the simulated random oracle to get \( H(t). \) If \( \text{coin}(t) = 0 \) then \( B \) halts, otherwise it forwards to \( \sigma_{i,t} = e(H(t)^{\nu}, w^2) e(g^{\nu \phi_t}, w^3) \).

### Challenge Phase

At the challenge phase \( A \) outputs a forger sum * , \( \sigma^*_t \) for a time interval \( t. \) \( B \) fetches the tuple \( R_i(t^\dagger) \) and:

- If \( \text{coin}(t^\dagger) = 0, \) then it aborts.

- Otherwise it solves the BCDH assumption by computing:

\[
I = \frac{(\sigma^*_t)}{e(g, \nu)^{2 \sum_{\nu}^*}} = e(H(t^\dagger), \nu)^{\nu \phi_k} e(g, \nu)^{\nu \sum_{\nu}^*} e(g, \nu)^{2 \sum_{\nu}^*} = e(H(t^\dagger), \nu)^{\nu \phi_k} e(g, \nu)^{2 \sum_{\nu}^*}
\]

Finally it outputs \( I = e(g, \nu)^{\nu \phi_k + \sum_{\nu}^*}, \) which is the solution to the BCDH problem.

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

**Lemma 2.** Let \( A \) be a probabilistic polynomial time adversary who colludes with a user \( U_k \) in the CRA-I scheme and outputs a Type-I forgery with non-negligible probability. Then, there is an adversary \( B \) that outputs a Type-I forgery for the BaseLine scheme with non-negligible probability.

**Proof.** \( B \) calls the \( O_{\text{Setup}}^{\text{CRA-I}} \) oracle which returns the public parameters \( pp = (H, e, g, g_2) \) and the secret key of the Aggregator \( sk_A. \) \( B \) relays this information to \( A. \) Whenever \( A \) calls the \( O_{\text{EncTag}}(t, \text{uid}, x_{i,t}) \) oracle, \( B \) in turn forwards the query to the \( O_{\text{EncTag}}(t, \text{uid}, x_{i,t}) \) oracle of the CRA-I game, which replies with \( c_{i,t} = H(t)^{ek_i} g^{\nu \phi_t}, \text{mtag}_{i,t} = [H(t)^{i} g^{\nu \phi_t}, w_i^2, t]^2. \) Similarly \( B \) relays the queries to the \( O_{\text{Tag}}(\text{mtag}_{i,t}) \) and forwards the response \( \sigma_{i,t} = e(H(t)^{\nu}, w)^2 e(g^{\nu \phi_t}, w)^2 \) back to \( A. \) \( B \) responds to the queries for \( O_{\text{Set}}^{\text{CRA-I}}(t, \text{uid} = i \in \mathbb{U}) \) oracle, with \( \nu, ek_i, \text{mtag}_{i,t}, w. \) Note the trusted server users \( A \) only learns \( H(t)^{ik} g^{\nu \phi_t}, [H(t)^{i} g^{\nu \phi_t}, w_i^2, t, e(H(t)^{i} g^{\nu \phi_t}, w)^2 \rangle \) by knowing \( w. \) Thus the secret value \( x_{i,t}, \) and the secret keys of the computation hidden. At this point the view of \( A \) is consistent with the real protocol and thus does not abort the game.

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

**Proof.** The proof proceeds accordingly with the previous proof for lemma 2. \( B \) relays queries to \( O_{\text{Setup}}^{\text{CRA-II}} \) oracles, coming from \( C. \) When \( C \) corrupts a user \( U_k \in \mathbb{S} \) then \( B \) forwards to \( C \) the secret keys \( (\nu, ek_i, tk_i, w). \) Finally the view of adversary \( C \) is identical with the real game without being able to distinguish since \( H(t)^{ik} g^{\nu \phi_t}, [H(t)^{i} g^{\nu \phi_t}, w_i^2, t, e(H(t)^{i} g^{\nu \phi_t}, w)^2 \rangle \) computationally hide the secret value \( x_{i,t} \) and \( (\nu, ek_i, tk_i) \) keys from uncorrupted users by an adversary \( C \) knowing the secret key \( r \) and secret keys of corrupted users.

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

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

**Proof.** (sketch) Notice the a CR – AU – II forgery entails collisions between a Converter and an Aggregator, by revealing \( r \) to the latter. Thus the proofs proceeds as with lemma 1, with the difference that \( A \) during the learning phase calls the \( O_{\text{Con}} \) oracle and \( B \) forwards to \( A \) the secret key \( r. \)

Due to space limitations the proof for Type-II forgeries is referred to the appendix section.

### 7.3 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 collision resistant unforgeability. The results are depicted in table 2. At each time in-
Obliviousness

Verifiability

Collusions

Shu et al. [23]  ✓  ✓  ✓

Joye et al. [15]  ✓  ✓  ✓

Erkin et al. [18]  ✓  ✓  ✓

Li et al. [10]  ✓  ✓  ✓

Jawurek et al. [22]  ✓  ✓  ✓

Kursawe et al. [17]  ✓  ✓  ✓

Barthe et al. [7]  ✓  ✓  ✓

Leontiadis et al. [28]  ✓  ✓  ✓

Leontiadis et al. [29]  ✓  ✓  ✓

Jung et al. [25]  ✓  ✓  ✓

This work  ✓  ✓  ✓

Table 3: Security comparison of existing protocols.

terval $t$, $U_i$ encrypts its data value $x_{i,t}$ with the secret encryption key $e_k$, as $c_{i,t} = H(t)^{g^{x_{i,t}}}$, thus resulting in two exponentiations and one hash evaluation in $G_1$. For the computation of the metatag $mtag_{i,t} = \left[H(t)^{g^{x_{i,t}}}, w^r \right]_G$, $U_i$ is committed to two exponentiations in $G_1$ and one exponentiation in $G_2$. Afterwards, in order to validate the final tag, users check if the following equation holds: $e(\sigma_{i,t}^2, w)^{\frac{1}{x_{i,t}}} = \sigma_{i,t}^2$, by performing one exponentiation in $G_T$ and one bilinear pairing operation. The Converter, in order to convert the metatag $mtag_{i,t}$ to the final tag $\sigma_{i,t} = e(H(t)^{g^{x_{i,t}}}, w)^e(g^{x_{i,t}}, w^r)$, is entitled in one bilinear pairing computation and one exponentiation in $G_T$. The Aggregator computes the proof with $n - 1$ multiplications in $G_T$: $\sigma_i = \prod_{t=1}^n \sigma_{i,t} = e(H(t), w)^{\sum_{t=1}^n r_i} e(g, w)^{r} \sum_{t=1}^n x_{i,t}$ and the data analyzer verifies with two multiplications in $G_T$, two exponentiations in $G_T$ and two bilinear pairing evaluations: $e(H(t), v_k)^{v_k \sum_{t=1}^n x_{i,t}} e(g, v_k)^{v_k}$, by Notice, that the protocol achieves constant verification time, which does not depend on the number of users that participate in the protocol with their values.

7.4 Comparison

We present a detailed comparison with respect to the security model and the collusion resistant property of existing protocols in Table 3. Protocols which assure Aggregator obliviousness (AO) protect individual privacy from semi-honest Aggregators. Interestingly, a recent published paper [25], necessitates the appropriate and rigorous security analysis that should be conducted for secure aggregation protocols. As already mentioned in [15] there are two flaws in [25]. 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 collusion, 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 [29] 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. [3] 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.

8. 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 in the random oracle model under standard assumptions and achieves constant time verification in the symmetric setting.

References:


APPENDIX

.1 Type-II Forgeries

**THEOREM 4.** An adversary \(\mathcal{A}\) who colludes with a user \(U_c\) in the CRA-II scheme has negligible probability on forging a Type-II \(\text{CR}^{\text{AU-I, II}}\) forgery, under the DL assumption in \(\mathbb{G}_2\) in the random oracle mode.

**Proof.** (sketch) For a \(\text{CR}^{\text{AU-I}}\) forgery an adversary observes metatags, tags and ciphertexts of users. In order to forge the proof \(\sigma_t = e(H(t), w)^{\sum_{i=1}^n r_i} e(g, w)^{\sum_{i=1}^n x_{i,t}}\), by colluding with a user \(U_c\), who reveals \(w, r_i\), \(\mathcal{A}\) should extract \(\sum_{i=1}^n r_i\) in order to compute \(e(H(t), w)^{\sum_{i=1}^n r_i}\). \(\mathcal{A}\) then recovers \(e(g, w)^{\sum_{i=1}^n x_{i,t}}\) and can forge by raising this element to a value \(v\) and multiplying again with \(\sigma_t = e(H(t), w)^{\sum_{i=1}^n r_i}\). Thus, it outputs as a valid forgery \((\text{sum}_t = v \sum x_{i,t}, \sigma_t = e(H(t), w)^{\sum_{i=1}^n r_i} e(g, w)^{v \sum_{i=1}^n x_{i,t}})\). But \(\sum_{i=1}^n r_i\) cannot be computed, since adversary cannot corrupt all users and is computationally hidden at the exponent.

Similarly the adversary for a \(\text{CR}^{\text{AU-II}}\) forgery has to extract the computationally hidden at the exponent \(\sum_{i=1}^n r_i\) by colluding with \(C\) and learning \(r\). \(\square\)