Secure Aggregation with Uncoded Groupwise Keys Against User Collusion

Ziting Zhang*, Kai Wan*, Hua Sun†, Mingyue Ji‡, Giuseppe Caire§
*Huazhong University of Science and Technology, 430074 Wuhan, China, {zzting,kai_wan}@hust.edu.cn
†University of North Texas, Denton, TX 76203, USA, hua.sun@unt.edu
‡University of Utah, Salt Lake City, UT 84112, USA, mingyue.ji@utah.edu
§Technische Universität Berlin, 10587 Berlin, Germany, caire@tu-berlin.de

2023 8th International Conference on Computer and Communication Systems (ICCCS) | 978-1-6654-5612-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCS57501.2023.10151414

***Abstract*—In this paper, we study the information theoretic secure aggregation problem, where the server node aims to aggregate K users' locally trained models, without revealing any other information about the users' local data. To ensure security, some keys are shared among the users, which is referred to as the key sharing phase. Uncoded groupwise keys are considered, where each key is shared by a subset of S users and is independent from other keys. After the key sharing phase, each user masks its trained model and sends to the server,which is referred to as the model aggregation phase. In the presence of users' dropouts (i.e., up to $\mathsf{K}-\mathsf{U}$ user may drop during the model aggregation phase and the identity of the dropped users cannot be predicted), to guarantee the information theoretic security, two-round transmissions are necessary. Our objective is to characterize the capacity region of the transmission rates (i.e., the normalized numbers of two-round transmissions by each user) in the two rounds. When $\mathsf{S}\geq\mathsf{K}-\mathsf{U}+1$, the capacity region was recently characterized. In this paper, we additionally consider the potential effect of user collusion, where there may exist up to T users colluding with the server. With the presence of the colluding users, the security constraint becomes that, except the sum of trained models, the server cannot learn any information about the other users' local data even if it colludes with any set of up to T users. For this new problem, we propose two secure aggregation schemes, which work for the cases of $\mathsf{S}=\mathsf{K}-\mathsf{U}+1$ and of $\mathsf{K}-\mathsf{U}+1\leq\mathsf{S}\leq\mathsf{K}-\mathsf{T}$, respectively. The first scheme is then proven to achieve the capacity region.**

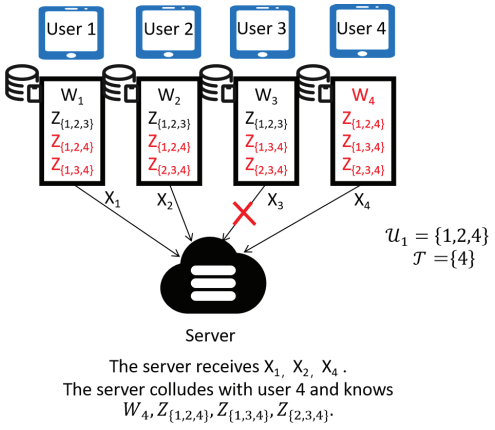
***Index Terms*—Secure aggregation; information theory; uncoded groupwise keys; user collusion**

I. Introduction

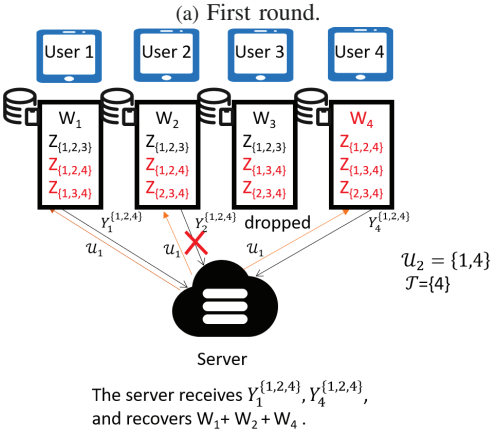
With the development of modern edge devices such as mobile phones, it is possible to access a large amount of data suitable for learning models. Federated learning (FL) leverages the edge devices' local data and computational resource to proceed trainings [1]. In a FL framework, users compute the trained model and send back the computation results to the central server; the server then updates the model with the received results aggregation [2]–[4]. Compared to other distributed machine learning scenarios, FL has a significant advantage in preserving the security of users' local data against the server, since the users do not need to transmit the original local data to the server.

To further guarantee that the server only gets the sum of updated models without retrieving any other information about the users' local data, secure aggregation for FL was originally introduced in [5], where various secure aggregation schemes were proposed with the tolerance against user dropout and collusion. Recently, an information theoretic $(\mathsf{K},\mathsf{U},\mathsf{T})$ secure aggregation problem against user dropout and collusion was proposed in [6]. The secure aggregation framework contains two phases. During the key sharing phase, the users share the keys in an offline scenario, where the generated keys are independent from the trained models in the future phase. During the model aggregation phase, each user first computes the trained model by using its local data. Assume that each trained model has L i.i.d. symbols on some finite field $\mathbb{F}_{qsf}$. To ensure security, a two-round transmission process is used. In the first round, each user sends a coded message to the server as a function of their trained model and shared key. Due to user dropout, the server only receives messages from the users in $\mathcal{U}_1$, where $\mathcal{U}_1\subseteq\{1,\ldots,\mathsf{K}\}$ and $|\mathcal{U}_1|\geq\mathsf{U}$. The server then informs the users in $\mathcal{U}_1$. In the second round, each user in $\mathcal{U}_1$ transmits linear combination of keys to the server. Due to user dropouts in the second round, the server receives answers from the users in $\mathcal{U}_2$, where $\mathcal{U}_2\subseteq\mathcal{U}_1$ and $|\mathcal{U}_2|\geq\mathsf{U}$. The decodability constraint is that the server should recover the sum of the trained models by the users in $\mathcal{U}_1$ from its received messages. For the security constraint, the server may collude with any subset of users $\mathcal{T}$ where $\mathcal{T}\subseteq\{1,\ldots,\mathsf{K}\}$ and $|\mathcal{T}|\leq\mathsf{T}$. Besides the sum of the trained models by the users in $\mathcal{U}_1$, the server should not know any other information about the trained models by the users in $\{1,\ldots,\mathsf{K}\}\setminus\mathcal{T}$ even if it knows the trained models and stored keys of the users in $\mathcal{T}$. According to [6], each user needs to send a minimum of L symbols in the first round and $\mathsf{L}/(\mathsf{U}-\mathsf{T})$ symbols in the second round, which can be achieved simultaneously.

In the secure aggregation schemes in [6], [7], which can achieve the minimum numbers of transmissions, the users store some coded keys. In [8], the information theoretic secure aggregation problem with uncoded groupwise keys was formulated. The keys shared by users are groupwise and uncoded, which is motivated by practical key generation techniques. Each key is stored by a set of users and is independent among each other. In contrast to the model proposed in [6], this formulation introduces an additional constraint on the uncoded groupwise keys, namely that each key is shared by S users.Without the consideration of user collusion, a secure aggregation scheme achieving the optimal communication rates was proposed in [8] when $\mathsf{S}\geq\mathsf{K}-\mathsf{U}+1$.



(a) First round.



(b) Second round.

Fig. 1: Information theoretic secure aggregation with uncoded groupwise keys: A case study with $(\mathsf{K},\mathsf{U},\mathsf{S}) = (4,2,3)$

Main Contributions: In this paper, we formulate the $(\mathsf{K},\mathsf{U},\mathsf{S},\mathsf{T})$ information theoretic secure aggregation problem with uncoded groupwise keys against user collusion, as illustrated in Fig. 1. For this new problem, we propose two secure aggregation schemes:

- The first scheme is from an extension from the secure aggregation scheme in [8] and works for the case $\mathsf{S} = \mathsf{K}-\mathsf{U}+1$. Interestingly, this scheme achieves the same communication rates as the optimal secure aggregation schemes [6], [7] which are built on coded keys.
- The second scheme works for the case $\mathsf{K}-\mathsf{U}+1 \leq \mathsf{S} \leq \mathsf{K}-\mathsf{T}$, and is built on a smart application of the coding strategy for distributed gradient descent [9]–[11].

The details of all proofs in this paper will be provided in our future extended version.

Notation Convention: Calligraphic symbols denote sets, bold symbols denote vectors and matrices, and sans-serif symbols denote system parameters. We use $|\cdot|$ to represent the cardinality of a set or the length of a vector; $[a:b] := \{a, a+1, \ldots, b\}$ and $[n] := [1:n]$; $\mathbb{F}_{\mathsf{q}}$ represents a finite field with order q; $\mathbf{M}^{\mathrm{T}}$ and $\mathbf{M}^{-1}$ represent the transpose and the inverse of matrix $\mathbf{M}$, respectively; the matrix $[a;b]$ is written in a Matlab form, representing $\begin{bmatrix} a \\ b \end{bmatrix}$; $\mathrm{rank}(\mathbf{M})$ represents the rank of matrix $\mathbf{M}$; we let $\binom{x}{y} = 0$ if $x < 0$ or $y < 0$ or $x < y$; for any set $\mathcal{S}$ and an integer s, we let $\binom{\mathcal{S}}{s}$ represent the collection of all subsets of $\mathcal{S}$ with s elements. Entropies will be in base q, where q represents the field size.

II. System Model

We formulate a $(\mathsf{K},\mathsf{U},\mathsf{S},\mathsf{T})$ information theoretic secure aggregation problem involving a server and $\mathsf{K} \geq 2$ users. Each user $k \in [\mathsf{K}]$ holds an input vector W_k and uncoded groupwise keys $Z_k = \left(Z_{\mathcal{V}} : \mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}, k \in \mathcal{V}\right)$, where $Z_{\mathcal{V}}$ is shared among S users. The input vectors in $(W_k : k \in [\mathsf{K}])$ are independent, consisting of L uniform and i.i.d. symbols over a finite field $\mathbb{F}_{\mathsf{q}}$. The keys in $\left(Z_{\mathcal{V}} : \mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}\right)$ are independent of each other and independent of the input vectors. Thus

$$H\left((W_k : k \in [\mathsf{K}]), (Z_{\mathcal{V}} : \mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}})\right) = \sum_{k \in [\mathsf{K}]} H(W_k) + \sum_{\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}} H(Z_{\mathcal{V}}). \tag{1}$$

The model aggregation phase contains two rounds of transmissions.

In the first round. User k sends the ciphertext X_k, which is a function of W_k and Z_k, to the server. In the first transmission round, a subset of users may drop and the surviving users are denoted by $\mathcal{U}_1$, where $\mathcal{U}_1 \subseteq [\mathsf{K}]$ and $|\mathcal{U}_1| \geq \mathsf{U}$; thus the server receives X_k where $k \in \mathcal{U}_1$. The communication rate for the first round is determined by the maximum transmission load of all users, i.e., $\mathsf{R}_1 := \max_{k \in [\mathsf{K}]} \frac{|X_k|}{\mathsf{L}}$.

In the second round. The server sends the value $\mathcal{U}_1$ back to the users in $\mathcal{U}_1$. User k then sends $Y_k^{\mathcal{U}_1}$, which is the linear combination of keys Z_k, to the server. In the second transmission round, a subset of users may still drop and the surviving users are denoted by $\mathcal{U}_2$, where $\mathcal{U}_2 \subseteq \mathcal{U}_1$ and $|\mathcal{U}_2| \geq \mathsf{U}$; thus the server receives $Y_k^{\mathcal{U}_1}$ where $k \in \mathcal{U}_2$. The communication rate for the second round is determined by the maximum transmission load of $\mathcal{U}_1$ users, i.e., $\mathsf{R}_2 := \max_{\mathcal{U}_1 \subseteq [\mathsf{K}]:|\mathcal{U}_1| \geq \mathsf{U}} \max_{k \in \mathcal{U}_1} \frac{|Y_k^{\mathcal{U}_1}|}{\mathsf{L}}$.

Decodability. The server can recover $\sum_{k \in \mathcal{U}_1} W_k$ from $(X_k : k \in \mathcal{U}_1)$ and $\left(Y_k^{\mathcal{U}_1} : k \in \mathcal{U}_2\right)$, i.e., for any $\mathcal{U}_1 \subseteq [\mathsf{K}]$ and $\mathcal{U}_2 \subseteq \mathcal{U}_1$,

$$H\left(\sum_{k \in \mathcal{U}_1} W_k \middle| (X_k : k \in \mathcal{U}_1), (Y_k^{\mathcal{U}_1} : k \in \mathcal{U}_2)\right) = 0. \tag{2}$$

Security. Even if the server may collude with any set of users $\mathcal{T}$ where $|\mathcal{T}| \leq T$, the server cannot get any information about the input vectors of the non-colluding users except $\sum_{k \in \mathcal{U}_1} W_k$. Thus for any $\mathcal{U}_1 \subseteq [\mathsf{K}]$ where $|\mathcal{U}_1| \geq \mathsf{U}$, (we assume the server can receive all possible transmissions,

i.e.,$(X_k : k \in [\mathsf{K}])$ and $(Y_k^{\mathcal{U}_1} : k \in \mathcal{U}_1)$)

$$I\big((W_k : k \in [K]); (X_k : k \in [\mathsf{K}]), (Y_k^{\mathcal{U}_1} : k \in \mathcal{U}_1)\Big| \sum_{k\in\mathcal{U}_1} W_k, (W_k, Z_k : k \in \mathcal{T})\big) = 0. \tag{3}$$

Note that in our problem, for the security constraint, we must have $\mathsf{T} < \mathsf{U}$ as shown in [6].

Objective. The rate tuple (R_1, R_2) is achievable if there exists a secure aggregation scheme satisfying (2) and (3), where the keys satisfy (1). Our goal is to determine the capacity region $\mathcal{R}^*$ (i.e., set of all achievable rate tuples).

A converse bound on the capacity region of our considered problem $(\mathsf{K}, \mathsf{U}, \mathsf{S}, \mathsf{T})$ can be obtained from the converse bound in [6], which is the converse bound for any possible key generations.

Lemma 1 ([6]). *For the* $(\mathsf{K}, \mathsf{U}, \mathsf{S}, \mathsf{T})$ *information theoretic secure aggregation problem where* $\mathsf{T} < \mathsf{U}$*, each achievable rate tuple* $(\mathsf{R}_1, \mathsf{R}_2)$ *satisfies*

$$\mathsf{R}_1 \geq 1,\ \mathsf{R}_2 \geq 1/(\mathsf{U} - \mathsf{T}). \tag{4}$$

When $\mathsf{T} = 0$, the following capacity results were characterized in [8].

Lemma 2 ([8]). *For the* $(\mathsf{K}, \mathsf{U}, \mathsf{S}, \mathsf{T})$ *information theoretic secure aggregation problem where* $\mathsf{T} = 0$ *and* $\mathsf{S} > \mathsf{K} - \mathsf{U}$*, every achievable rate tuple* $(\mathsf{R}_1, \mathsf{R}_2)$ *satisfies*

$$\mathsf{R}_1 \geq 1,\ \mathsf{R}_2 \geq 1/\mathsf{U}. \tag{5}$$

III. Main Results

Theorem 1. *For the* $(\mathsf{K}, \mathsf{U}, \mathsf{S}, \mathsf{T})$ *information theoretic secure aggregation problem where* $\mathsf{S} = \mathsf{K} - \mathsf{U} + 1$ *and* $\mathsf{T} < \mathsf{U}$*,*

$$\mathcal{R}^\star = \{(\mathsf{R}_1, \mathsf{R}_2) : \mathsf{R}_1 \geq 1, \mathsf{R}_2 \geq 1/(\mathsf{U} - \mathsf{T})\}. \tag{6}$$

The converse bound for Theorem 1 follows directly from Lemma 1, and the proposed achievable scheme is described in Section IV.

In [8], the number of keys needed is K when $\mathsf{U} \leq \mathsf{K} - \mathsf{U} + 1$ and is $\mathcal{O}(\mathsf{K}^2)$ when $\mathsf{U} > \mathsf{K} - \mathsf{U} + 1$. But if there are colluding users, this number of keys will not be enough to guarantee the security constraint. Instead, we propose to use all $\binom{\mathsf{K}}{\mathsf{S}}$ keys in our secure scheme.

The transmission rates of the proposed secure aggregation scheme in Theorem 1 are exactly the same as the optimal secure aggregation scheme in [6]. Hence, in the case of $\mathsf{S} = \mathsf{K} - \mathsf{U} + 1$, the uncoded groupwise keys can achieve the general optimality among all possible key generations. Next, we propose our second secure aggregation scheme for the more general case where $\mathsf{K} - \mathsf{U} + 1 \leq \mathsf{S} \leq \mathsf{K} - \mathsf{T}$.

Theorem 2. *For the* $(\mathsf{K}, \mathsf{U}, \mathsf{S}, \mathsf{T})$ *information theoretic secure aggregation problem where* $\mathsf{K} - \mathsf{U} + 1 \leq \mathsf{S} \leq \mathsf{K} - \mathsf{T}$ *and* $\mathsf{T} < \mathsf{U}$*, the following rate region is achievable*

$$\mathsf{R}_1 \geq 1, \mathsf{R}_2 \geq \frac{1}{\mathsf{S} + \mathsf{U} - \mathsf{K}}. \tag{7}$$

IV. Proof of Theorem 1: achievability

For each set $\mathcal{V}$ where $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}$, we choose one vector $\mathbf{a}_\mathcal{V} = [a_{\mathcal{V},1}, \ldots, a_{\mathcal{V},\mathsf{U}}]^\mathsf{T}$, where each element $a_{\mathcal{V},j}$ is a coefficient in $\mathbb{F}_\mathsf{q}$ to be designed later.

In the first round, for each $k \in [\mathsf{K}]$ we divide W_k into $\mathsf{U} - \mathsf{T}$ pieces, each with $\mathsf{L}/(\mathsf{U} - \mathsf{T})$ uniformly i.i.d. symbols over $\mathbb{F}_\mathsf{q}$. Each user k sends

$$X_{k,j} = W_{k,j} + \sum_{\mathcal{V}\in\binom{[\mathsf{K}]}{\mathsf{S}}: k\in\mathcal{V}} a_{\mathcal{V},j} Z_{\mathcal{V},k},\ \forall j \in [\mathsf{U} - \mathsf{T}], \tag{8}$$

where $X_{k,j}$ contains $\mathsf{L}/(\mathsf{U} - \mathsf{T})$ symbols, and the coefficients $a_{\mathcal{V},j} \in \mathbb{F}_\mathsf{q}$ are designed accordingly. The vector $X_k = (X_{k,1}, \ldots, X_{k,U-T})$ contains L symbols, resulting in $R_1 = 1$. In the first round, the server receives X_k for each user $k \in \mathcal{U}_1$,and thus recovers

$$\begin{aligned}\sum_{k\in\mathcal{U}_1} X_{k,j} &= \sum_{k\in\mathcal{U}_1} W_{k,j} + \sum_{\mathcal{V}\in\binom{[\mathsf{K}]}{\mathsf{S}}: \mathcal{V}\cap\mathcal{U}_1\neq\emptyset} \left(a_{\mathcal{V},j} \sum_{k_1\in\mathcal{V}\cap\mathcal{U}_1} Z_{\mathcal{V},k_1}\right) \\ &= \sum_{k\in\mathcal{U}_1} W_{k,j} + \sum_{\mathcal{V}\in\binom{[\mathsf{K}]}{\mathsf{S}}} \left(a_{\mathcal{V},j} \sum_{k_1\in\mathcal{V}\cap\mathcal{U}_1} Z_{\mathcal{V},k_1}\right),\ \forall j \in [\mathsf{U} - \mathsf{T}],\end{aligned} \tag{9}$$

where (9) follows since $\mathsf{S} = \mathsf{K} - \mathsf{U} + 1$ and thus there is no $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}$ where $\mathcal{V} \cap \mathcal{U}_1 = \emptyset$. It can be seen from (9) that the server still needs to recover $\sum_{\mathcal{V}\in\binom{[\mathsf{K}]}{\mathsf{S}}} \left(a_{\mathcal{V},j} \sum_{k_1\in\mathcal{V}\cap\mathcal{U}_1} Z_{\mathcal{V},k_1}\right)$ for each $j \in [\mathsf{U} - \mathsf{T}]$ in the next round. For the sake of ease notation, we define

$$Z_\mathcal{V}^{\mathcal{U}_1} := \sum_{k_1\in\mathcal{V}\cap\mathcal{U}_1} Z_{\mathcal{V},k_1},\ \forall \mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}, \tag{10}$$

This quantity contains $\mathsf{L}/(\mathsf{U} - \mathsf{T})$ uniform and i.i.d. symbols. By the construction of the first round transmission, the server only needs to further recover $\sum_{\mathcal{V}\in\binom{[\mathsf{K}]}{\mathsf{S}}} a_{\mathcal{V},j} Z_\mathcal{V}^{\mathcal{U}_1}$ for each $j \in [\mathsf{U}]$ in the second round to obtain $\sum_{k\in\mathcal{U}_1} W_{k,j}$.

In the second round, we denote the sets in $\binom{[\mathsf{K}]}{\mathsf{S}}$ by $\mathcal{S}(1), \ldots, \mathcal{S}\left(\binom{\mathsf{K}}{\mathsf{S}}\right)$, and for each $k \in [K]$ denote the sets in $\binom{[\mathsf{K}]\setminus\{k\}}{\mathsf{S}}$ by $\mathcal{S}_{\overline{k}}(1), \ldots, \mathcal{S}_{\overline{k}}\left(\binom{\mathsf{K}-1}{\mathsf{S}}\right)$. We let the server recover

$$\begin{bmatrix} F_1 \\ \vdots \\ F_U \end{bmatrix} = \left[\mathbf{a}_{\mathcal{S}(1)}, \ldots, \mathbf{a}_{\mathcal{S}\left(\binom{\mathsf{K}}{\mathsf{S}}\right)}\right] \begin{bmatrix} Z_{\mathcal{S}(1)}^{\mathcal{U}_1} \\ \vdots \\ Z_{\mathcal{S}\left(\binom{\mathsf{K}}{\mathsf{S}}\right)}^{\mathcal{U}_1} \end{bmatrix}, \tag{11}$$

where each F_j, $j \in [\mathsf{U}]$, contains $\mathsf{L}/(\mathsf{U} - \mathsf{T})$ symbols.

Each user $k \in \mathcal{U}_1$ sends

$$Y_k^{\mathcal{U}_1} = \mathbf{s}_k \begin{bmatrix} F_1 \\ \vdots \\ F_\mathsf{U} \end{bmatrix}, \tag{12}$$

where $\mathbf{s}_k$ represents a left null space vector of $\left[\mathbf{a}_{\mathcal{S}_{\overline{k}}(1)}, \ldots, \mathbf{a}_{\mathcal{S}_{\overline{k}}\left(\binom{\mathsf{K}-1}{\mathsf{S}}\right)}\right]$. Note that $Y_k^{\mathcal{U}_1}$ contains $\mathsf{L}/(\mathsf{U} - \mathsf{T})$ symbols, leading to $\mathsf{R}_2 = 1/(\mathsf{U} - \mathsf{T})$.

For each $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}$, we define $\mathbf{a}'_{\mathcal{V}}$ as the sub-vector including the first $\mathsf{U}-\mathsf{T}$ elements of $\mathbf{a}_{\mathcal{V}}$. For each $\mathcal{T} \subseteq [\mathsf{K}]$ where $|\mathcal{T}| \leq \mathsf{T}$ and each $k \in [\mathsf{K}] \backslash \mathcal{T}$, we sort all sets $\mathcal{S} \in \binom{[\mathsf{K}]}{\mathsf{S}}$ where $k \in \mathcal{S}$ and $\mathcal{S} \cap \mathcal{T} = \emptyset$ in a lexicographic order and denote them by $\mathcal{S}_{k,\overline{\mathcal{T}}}(1), \ldots, \mathcal{S}_{k,\overline{\mathcal{T}}}\left(\binom{\mathsf{K}-|\mathcal{T}|-1}{\mathsf{S}-1}\right)$. To satisfy the decodability and security constraints, our choose of the coefficient vectors $\mathbf{a}_{\mathcal{V}}$ where $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}$ has the following properties.

Property 1. *For each $k \in [\mathsf{K}]$ and each $\mathcal{T} \subseteq [\mathsf{K}] \setminus \{k\}$ where $|\mathcal{T}| \leq \mathsf{T}$,*

$$\left[\mathbf{a}'_{\mathcal{S}_{k,\overline{\mathcal{T}}}(1)}, \ldots, \mathbf{a}'_{\mathcal{S}_{k,\overline{\mathcal{T}}}\left(\binom{\mathsf{K}-|\mathcal{T}|-1}{\mathsf{S}-1}\right)}\right] \quad \textit{has rank equal to } \mathsf{U}-\mathsf{T}. \tag{13}$$

Property 1 guarantees that even if the server colludes with T users, it cannot learn any information about W_k from X_k.

Property 2. *For each user $k \in [\mathsf{K}]$,*

$$\left[\mathbf{a}_{\mathcal{S}_{\overline{k}}(1)}, \ldots, \mathbf{a}_{\mathcal{S}_{\overline{k}}\left(\binom{\mathsf{K}-1}{\mathsf{S}}\right)}\right] \quad \textit{has rank equal to } \mathsf{U}-1. \tag{14}$$

Property 2 guarantees that the left null space of $\left[\mathbf{a}_{\mathcal{S}_{\overline{k}}(1)}, \ldots, \mathbf{a}_{\mathcal{S}_{\overline{k}}\left(\binom{\mathsf{K}-1}{\mathsf{S}}\right)}\right]$ exists, such that the encodability of user k in the second round transmission is guaranteed.

Property 3.

$$\textit{Any } \mathsf{U} \textit{ vectors in } \{\mathbf{s}_k : k \in \mathcal{U}_1\} \textit{ are linearly independent.} \tag{15}$$

From any set of surviving users in the second round $\mathcal{U}_2 \subseteq \mathcal{U}_1$ where $|\mathcal{U}_2| \geq \mathsf{U}$, the server should recover $F_1, \ldots, F_{\mathsf{U}}$. So Property 3 guarantees that the server can recover $F_1, \ldots, F_{\mathsf{U}}$ in the second round, from the answers of any set of users $\mathcal{U}_2$.

Property 4. *For any $\mathcal{T} \subseteq [\mathsf{K}]$ where $|\mathcal{T}| \leq \mathsf{T}$, by denoting all sets $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}$ where $\mathcal{V} \cap \mathcal{T} = \emptyset$ by $\mathcal{S}_{\overline{\mathcal{T}}}(1), \ldots, \mathcal{S}_{\overline{\mathcal{T}}}\left(\binom{\mathsf{K}-\mathsf{T}}{\mathsf{S}}\right)$,*

$$\left[\mathbf{a}_{\mathcal{S}_{\overline{\mathcal{T}}}(1)}, \ldots, \mathbf{a}_{\mathcal{S}_{\overline{\mathcal{T}}}\left(\binom{\mathsf{K}-\mathsf{T}}{\mathsf{S}}\right)}\right] \quad \textit{has rank equal to } \mathsf{U}-|\mathcal{T}|. \tag{16}$$

On the condition of secure transmissions in the first round, Property 4 guarantees the second round transmission does not hurts the security constraint neither.

Due to the limitation of pages, we skip the general description on the choice of the above coefficient vectors, as well as the proofs of decodability and security for the proposed scheme. Instead, we will illustrate the main ingredients through the following example.

Example 1 $((\mathsf{K},\mathsf{U},\mathsf{S},\mathsf{T})) = (6,4,3,1))$**.** The proposed scheme is inspired from the secure aggregation scheme in [8] for the case $\mathsf{T}=0$. In this example, since $\mathsf{T}=1$, the server may collude with one or zero user. As explained in [6], without loss of generality, we can assume that q is large enough. In this example we further assume that q is a large prime number, which is not necessary in our general scheme.

We divide each input vector W_k, where $k \in [6]$, into $\mathsf{U}-\mathsf{T} = 3$ pieces $W_k = (W_{k,1}, W_{k,2}, W_{k,3})$. For each $\mathcal{V} \in \binom{[6]}{3}$, we generate $Z_{\mathcal{V}} = (Z_{\mathcal{V},k} : \mathsf{k} \in \mathcal{V})$ shared by all users in $\mathcal{V}$, where each $Z_{\mathcal{V},k}$ contains $\mathsf{L}/3$ uniformly i.i.d. symbols over $\mathbb{F}_{\mathsf{q}}$.

The next step is to select the U-dimensional coefficient vectors $\mathbf{a}_{\mathcal{V}}$ where $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}}$.

First, we select $\mathbf{a}_{\{1,2,3\}}$, $\mathbf{a}_{\{1,2,4\}}$, $\mathbf{a}_{\{1,2,5\}}$, $\mathbf{a}_{\{1,2,6\}}$ as the basis vectors, where the other vectors are located at the linear space spanned by the basis vectors. We select $[\mathbf{a}_{\{1,2,3\}}, \mathbf{a}_{\{1,2,4\}}, \mathbf{a}_{\{1,2,5\}}, \mathbf{a}_{\{1,2,6\}}]$ as a 4×4 Minimum Distance Separable (MDS) matrix; one possibility could be $[\mathbf{a}_{\{1,2,3\}}, \mathbf{a}_{\{1,2,4\}}, \mathbf{a}_{\{1,2,5\}}, \mathbf{a}_{\{1,2,6\}}] = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 4 & 2 & 5 & 3 \\ 3 & 1 & 7 & 1 \\ 5 & 3 & 1 & 0 \end{bmatrix} = [\mathsf{m}_{4,1}, \mathsf{m}_{4,2}, \mathsf{m}_{4,3}, \mathsf{m}_{4,4}]$. Let us define $\mathcal{G}_1 = \{\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,2,6\}\}$. By determining the basis vectors $\mathbf{a}_{\mathcal{V}'}$ where $\mathcal{V}' \in \mathcal{G}_1$, for each $\mathcal{V} \in \binom{[6]}{3} \setminus \mathcal{G}_1$, we look for the minimum subset of $\mathcal{G}_1$ the union of whose elements are a superset of $\mathcal{V}$, and then let $\mathbf{a}_{\mathcal{V}}$ be a linear combination of $\mathbf{a}_{\mathcal{V}_1}$ where $\mathcal{V}_1$ is in the found set. Next we determine coefficients in the linear combination.

We consider each $\mathcal{V} \in \binom{[\mathsf{K}]}{\mathsf{S}} \setminus \mathcal{G}_1$ where $\{3,4\} \subseteq \mathcal{V}$. For example, when $\mathcal{V} = \{1,3,4\}$, the minimum subset of $\mathcal{G}_1$ the union of whose elements are a superset of $\{1,3,4\}$ is $\{\{1,2,3\},\{1,2,4\}\}$; we let $\mathbf{a}_{\{1,3,4\}}$ be a random linear combination of $\mathbf{a}_{\{1,2,3\}}$ and $\mathbf{a}_{\{1,2,4\}}$,

$$\mathbf{a}_{\{1,3,4\}} = \mathbf{a}_{\{1,2,3\}} + 4\mathbf{a}_{\{1,2,4\}} = \mathbf{m}_{4,1} + 4\mathbf{m}_{4,2}.$$

Similarly, we have

$$\begin{aligned}
\mathbf{a}_{\{2,3,4\}} &= \mathbf{a}_{\{1,2,3\}} + 8\mathbf{a}_{\{1,2,4\}} = \mathbf{m}_{4,1} + 8\mathbf{m}_{4,2}, \\
\mathbf{a}_{\{3,4,5\}} &= \mathbf{a}_{\{1,2,3\}} + \mathbf{a}_{\{1,2,4\}} + \mathbf{a}_{\{1,2,5\}} \\
&= \mathbf{m}_{4,1} + \mathbf{m}_{4,2} + \mathbf{m}_{4,3}, \\
\mathbf{a}_{\{3,4,6\}} &= \mathbf{a}_{\{1,2,3\}} + 2\mathbf{a}_{\{1,2,4\}} + \mathbf{a}_{\{1,2,6\}} \\
&= \mathbf{m}_{4,1} + 2\mathbf{m}_{4,2} + \mathbf{m}_{4,4}.
\end{aligned}$$

Define $\mathcal{G}_2 = \{\{1,3,4\},\{2,3,4\},\{3,4,5\},\{3,4,6\}\}$.

For each set $\mathcal{V} \in \binom{[6]}{3} \backslash (\mathcal{G}_1 \cup \mathcal{G}_2)$, we search for the minimal subset of $\mathcal{G}_2$ whose union of whose elements is a superset of $\mathcal{V}$. Assume that this set is $\mathcal{F}$. Let $\mathbf{a}_{\mathcal{V}}$ be a linear combination of $\mathbf{a}_{\mathcal{V}_1}$ where $\mathcal{V}_1 \in \mathcal{F}$.

If $\mathcal{V} \in \binom{[6]}{3} \setminus (\mathcal{G}_1 \cup \mathcal{G}_2)$ where $3 \in \mathcal{V}$, for example $\mathcal{V} = \{1,3,5\}$ $\mathbf{a}_{\{1,3,5\}}$ is also a linear combination of $\mathbf{a}_{\{1,2,3\}}$ and $\mathbf{a}_{\{1,2,5\}}$; thus $\mathbf{a}_{\{1,3,5\}}$ does not contains $\mathbf{m}_{4,2}$ and only has a unique linear combination representation of $\mathbf{a}_{\{1,3,4\}}$ and $\mathbf{a}_{\{3,4,5\}}$, which is

$$\mathbf{a}_{\{1,3,5\}} = -\mathbf{a}_{\{1,3,4\}} + 4\mathbf{a}_{\{3,4,5\}} = 3\mathbf{m}_{4,1} + 4\mathbf{m}_{4,3}.$$

Similarly, we can fix $\mathbf{a}_{\{1,3,6\}}$, $\mathbf{a}_{\{2,3,5\}}$, $\mathbf{a}_{\{2,3,6\}}$, and $\mathbf{a}_{\{3,5,6\}}$ in Table I.

If $\mathcal{V} \in \binom{[6]}{3} \setminus (\mathcal{G}_1 \cup \mathcal{G}_2)$ where $4 \in \mathcal{V}$, $\mathbf{a}_{\mathcal{V}}$ does not contains $\mathbf{m}_{4,1}$ and only has a unique linear combination representation. For example, $\mathbf{a}_{\{1,4,5\}}$ is a linear combination of $\mathbf{a}_{\{1,3,4\}}$ and $\mathbf{a}_{\{3,4,5\}}$, and does not contain $\mathbf{m}_{4,1}$. Thus we have

$$\mathbf{a}_{\{1,4,5\}} = \mathbf{a}_{\{1,3,4\}} - \mathbf{a}_{\{3,4,5\}} = 3\mathbf{m}_{4,2} - \mathbf{m}_{4,3}.$$

TABLE I: Coefficient vectors $\mathbf{a}_{\mathcal{V}}$ in the $(\mathsf{K},\mathsf{U},\mathsf{S},\mathsf{T})=(6,4,3,1)$ information theoretic secure aggregation problem.

$\mathbf{a}_{\mathcal{V}}$	Composition	Value	$\mathbf{a}_{\mathcal{V}}$	Composition	Value
$\mathbf{a}_{\{1,2,3\}}$	$\mathbf{m}_{4,1}$	$\mathbf{m}_{4,1}$	$\mathbf{a}_{\{2,3,4\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,2}$	$\mathbf{m}_{4,1}+8\mathbf{m}_{4,2}$
$\mathbf{a}_{\{1,2,4\}}$	$\mathbf{m}_{4,2}$	$\mathbf{m}_{4,2}$	$\mathbf{a}_{\{2,3,5\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,3}$	$7\mathbf{m}_{4,1}+8\mathbf{m}_{4,3}$
$\mathbf{a}_{\{1,2,5\}}$	$\mathbf{m}_{4,3}$	$\mathbf{m}_{4,3}$	$\mathbf{a}_{\{2,3,6\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,4}$	$3\mathbf{m}_{4,1}+4\mathbf{m}_{4,4}$
$\mathbf{a}_{\{1,2,6\}}$	$\mathbf{m}_{4,4}$	$\mathbf{m}_{4,4}$	$\mathbf{a}_{\{2,4,5\}}$	$\mathbf{m}_{4,2},\mathbf{m}_{4,3}$	$7\mathbf{m}_{4,2}-\mathbf{m}_{4,3}$
$\mathbf{a}_{\{1,3,4\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,2}$	$\mathbf{m}_{4,1}+4\mathbf{m}_{4,2}$	$\mathbf{m}_{\{2,4,6\}}$	$\mathbf{m}_{4,2},\mathbf{m}_{4,4}$	$6\mathbf{m}_{4,2}-\mathbf{m}_{4,4}$
$\mathbf{a}_{\{1,3,5\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,3}$	$3\mathbf{m}_{4,1}+4\mathbf{m}_{4,3}$	$\mathbf{a}_{\{2,5,6\}}$	$\mathbf{m}_{4,3},\mathbf{m}_{4,4}$	$6\mathbf{m}_{4,3}-7\mathbf{m}_{4,4}$
$\mathbf{a}_{\{1,3,6\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,4}$	$\mathbf{m}_{4,1}+2\mathbf{m}_{4,4}$	$\mathbf{a}_{\{3,4,5\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,2},\mathbf{m}_{4,3}$	$\mathbf{m}_{4,1}+\mathbf{m}_{4,2}+\mathbf{m}_{4,3}$
$\mathbf{a}_{\{1,4,5\}}$	$\mathbf{m}_{4,2},\mathbf{m}_{4,3}$	$3\mathbf{m}_{4,2}-\mathbf{m}_{4,3}$	$\mathbf{a}_{\{3,4,6\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,2},\mathbf{m}_{4,4}$	$\mathbf{m}_{4,1}+2\mathbf{m}_{4,2}+\mathbf{m}_{4,4}$
$\mathbf{a}_{\{1,4,6\}}$	$\mathbf{m}_{4,2},\mathbf{m}_{4,4}$	$2\mathbf{m}_{4,2}-\mathbf{m}_{4,4}$	$\mathbf{a}_{\{3,5,6\}}$	$\mathbf{m}_{4,1},\mathbf{m}_{4,3},\mathbf{m}_{4,4}$	$\mathbf{m}_{4,1}+2\mathbf{m}_{4,3}-\mathbf{m}_{4,4}$
$\mathbf{a}_{\{1,5,6\}}$	$\mathbf{m}_{4,3},\mathbf{m}_{4,4}$	$2\mathbf{m}_{4,3}-3\mathbf{m}_{4,4}$	$\mathbf{a}_{\{4,5,6\}}$	$\mathbf{m}_{4,2},\mathbf{m}_{4,3},\mathbf{m}_{4,4}$	$-\mathbf{m}_{4,2}+\mathbf{m}_{4,3}-\mathbf{m}_{4,4}$

Similarly, we can fix $\mathbf{a}_{\{1,4,6\}}$, $\mathbf{a}_{\{2,4,5\}}$, $\mathbf{a}_{\{2,4,6\}}$, and $\mathbf{a}_{\{4,5,6\}}$ in Table I.

For $\mathcal{V}\in\binom{[6]}{3}\setminus(\mathcal{G}_1\cup\mathcal{G}_2)$ where $3,4\notin\mathcal{V}$, we let

$$\begin{aligned}\mathbf{a}_{\{1,5,6\}}&=-2\mathbf{a}_{\{1,4,5\}}+3\mathbf{a}_{\{1,4,6\}}\\&=1/2\mathbf{a}_{\{1,3,5\}}-3/2\mathbf{a}_{\{1,3,6\}}=2\mathbf{m}_{4,3}-3\mathbf{m}_{4,4},\\\mathbf{a}_{\{2,5,6\}}&=-6\mathbf{a}_{\{2,4,5\}}+7\mathbf{a}_{\{2,4,6\}}\\&=3/4\mathbf{a}_{\{2,3,5\}}-7/4\mathbf{a}_{\{2,3,6\}}=6\mathbf{m}_{4,3}-7\mathbf{m}_{4,4}.\end{aligned}$$

Thus we have determined each $\mathbf{a}_{\mathcal{V}}$ where $\mathcal{V}\in\binom{[\mathsf{K}]}{\mathsf{S}}$ as shown in Table I. Note that in the secure aggregation scheme [8], the vectors $\mathbf{a}_{\{1,4,5\}}$, $\mathbf{a}_{\{1,4,6\}}$, $\mathbf{a}_{\{1,5,6\}}$, $\mathbf{a}_{\{2,4,5\}}$, $\mathbf{a}_{\{2,4,6\}}$, $\mathbf{a}_{\{2,5,6\}}$, $\mathbf{a}_{\{4,5,6\}}$ are all zero vectors.

We have the following definitions for ease of notation, $\left[\mathbf{a}'_{\{1,2,3\}},\mathbf{a}'_{\{1,2,4\}},\ldots,\mathbf{a}'_{\{1,5,6\}},\ldots,\mathbf{a}'_{\{4,5,6\}}\right]$ as $\mathbf{S}'_1$, $\left[\mathbf{a}'_{\{1,2,3\}},\mathbf{a}'_{\{1,2,4\}},\ldots,\mathbf{a}'_{\{1,5,6\}}\right]$ as $\mathbf{S}'_2$ and $\left[\mathbf{a}_{\{1,2,3\}},\mathbf{a}_{\{1,2,4\}},\ldots,\mathbf{a}_{\{1,5,6\}},\ldots,\mathbf{a}_{\{4,5,6\}}\right]$ as $\mathbf{S}_1$, $\left[\mathbf{a}_{\{2,3,4\}},\mathbf{a}_{\{2,3,5\}},\ldots,\mathbf{a}_{\{4,5,6\}}\right]$ as $\mathbf{S}_2$.

First round. Each user sends the information protected by keys to the server. Since we divide the input into 3 pieces, we only need to use the first three rows of the coefficient matrix.

User 1 sends $X_1=(X_{1,1},X_{1,2},X_{1,3})$, where

$$\begin{bmatrix}X_{1,1}\\X_{1,2}\\X_{1,3}\end{bmatrix}=\begin{bmatrix}W_{1,1}\\W_{1,2}\\W_{1,3}\end{bmatrix}+\mathbf{S}'_2\begin{bmatrix}Z_{\{1,2,3\},1}\\\vdots\\Z_{\{1,5,6\},1}\end{bmatrix}.$$

Similarly, the transmissions by each user $k\in[6]$ can be described as $X_{k,j}=W_{k,j}+\sum_{\mathcal{V}\in\binom{[6]}{3}:k\in\mathcal{V}}a_{\mathcal{V},j}Z_{\mathcal{V},k}$, $\forall j\in[3]$.

In the first round, the server receives the transmissions by user $k\in\mathcal{U}_1$, and recovers (recall that $Z_{\mathcal{V}}^{\mathcal{U}_1}$ is defined in (10))

$$\begin{bmatrix}\sum_{k\in\mathcal{U}_1}X_{k,1}\\\sum_{k\in\mathcal{U}_1}X_{k,2}\\\sum_{k\in[\mathcal{U}_1}X_{k,3}\end{bmatrix}=\begin{bmatrix}\sum_{k\in\mathcal{U}_1}W_{k,1}\\\sum_{k\in\mathcal{U}_1}W_{k,1}\\\sum_{k\in\mathcal{U}_1}w_{k,1}\end{bmatrix}+\mathbf{S}'_1\begin{bmatrix}Z_{\{1,2,3\}}^{\mathcal{U}_1}\\\vdots\\Z_{\{4,5,6\}}^{\mathcal{U}_1}\end{bmatrix}.$$

In X_1, if the server colludes with user 4 and knows $Z_{\{1,2,4\},1},Z_{\{1,3,4\},1},Z_{\{1,4,5\},1},Z_{\{1,4,6\},1},W_1$ is perfectly protected by $(Z_{\{1,2,3\},1},Z_{\{1,2,5\},1},Z_{\{1,2,6\},1})$, since the submatrix of $\mathbf{S}'_1$ including the columns corresponding to $(Z_{\{1,2,3\},1},Z_{\{1,2,5\},1},Z_{\{1,2,6\},1})$ has equal to 3. Hence, the server cannot get any information about W_1 from X_1. Similarly, the transmissions in the first round are secure.

Second round. In order to recover $\sum_{k\in\mathcal{U}_1}W_k$, in the second round we let the server recover

$$\begin{bmatrix}F_1\\F_2\\F_3\\F_4\end{bmatrix}=\mathbf{S}_1\begin{bmatrix}Z_{\{1,2,3\}}^{\mathcal{U}_1}\\\vdots\\Z_{\{4,5,6\}}^{\mathcal{U}_1}\end{bmatrix}.\tag{17}$$

We let each user $k\in\mathcal{U}_2$ transmit in the second round a linear combination of $F_1,\ldots,F_4$. Assume that $1\in\mathcal{U}_1$. User 1 cannot encode $Z_{\mathcal{V}}^{\mathcal{U}_1}$ where $\mathcal{V}\in\binom{[2:6]}{3}$. By the choice of coefficient vectors in Table I, the matrix $\mathbf{S}_2$ has rank equal to 3, which is the same as the rank of $[\mathbf{a}_{\{2,3,4\}},\mathbf{a}_{\{3,4,5\}},\mathbf{a}_{\{3,4,6\}}]$. Thus the left null space of $\mathbf{S}_2$ contains exactly one linearly independent vector, which could be $[-497,-137,134,335]$. So we let user 1 send

$$Y_1^{\mathcal{U}_1}=-497F_1-137F_2+134F_3+335F_4.$$

Similarly, if user $k\in\mathcal{U}_1\setminus\{1\}$, user k sends $Y_k^{\mathcal{U}_1}$, where

$$\begin{aligned}Y_2^{\mathcal{U}_1}&=-95F_1-3F_2+18F_3+45F_4,\\Y_3^{\mathcal{U}_1}&=-47F_1+13F_2+8F_3+20F_4,\\Y_4^{\mathcal{U}_1}&=25F_1-6F_2-7F_3+4F_4,\\Y_5^{\mathcal{U}_1}&=-10F_1+11F_2-23F_3+7F_4,\\Y_6^{\mathcal{U}_1}&=4F_1-56F_2-35F_3+23F_4.\end{aligned}$$

By construction, for any $\mathcal{U}_2\subseteq\mathcal{U}_1$ where $|\mathcal{U}_2|=4$, $(Y_k^{\mathcal{U}_1}:k\in\mathcal{U}_2)$ are linearly independent; therefore, the server can recover F_1,F_2,F_3,F_4, and then recover $\sum_{k\in\mathcal{U}_1}W_k$.

Next we check the security of the proposed scheme. Let us assume $\mathcal{U}_1=[6]$. First, we consider the case where $|\mathcal{T}|=1$; for example $\mathcal{T}=\{4\}$. By colluding with user 4, the server knows $(Z_{\mathcal{V}}:4\in\mathcal{V})$. The coefficient matrix corresponding to $\mathbf{a}_{\mathcal{V}}$ where $4\notin\mathcal{V}$, $[\mathbf{a}_{\{1,2,3\}},\mathbf{a}_{\{1,2,5\}},\mathbf{a}_{\{1,2,6\}},\mathbf{a}_{\{1,3,5\}},\ldots,\mathbf{a}_{\{3,5,6\}}]$ has rank equal to 3, which is the same as the rank of $[\mathbf{a}_{\{1,2,3\}},\mathbf{a}_{\{1,2,5\}},\mathbf{a}_{\{1,2,6\}}]$. Intuitively, from the first round transmissions, the server cannot get any information about the input vectors except W_4. From the second round transmissions, the server can get 3 linear combinations of keys,

and then can obtain at most 3 linear combinations of $W_{k,j}$'s. By the decodability, the server can obtain $\sum_{k\in\mathcal{U}_1} W_{k,j}$ where $j \in [3]$ by the help of the second round transmissions. Hence, the server cannot get any other information about the input vectors, and thus the proposed scheme is secure in this case.

We then consider the case that $\mathcal{T} = \emptyset$, and prove the security of the proposed scheme by using a genie-aided method. Assume that in the first round, we generate $W_{k,4}$ for each $k \in [6]$, where $W_{k,4}$ contains $\mathsf{L}/3$ uniformly i.i.d. symbols over $\mathbb{F}_\mathsf{q}$. For each $k \in [6]$, we let $W'_k = (W_{k,1}, \ldots, W_{k,4})$. Assume that besides $X_{k,1}, X_{k,2}, X_{k,3}$, in the first round user k also transmits $X_{k,4} = W_{k,4} + \sum_{\mathcal{V}\in\binom{[6]}{3}:k\in\mathcal{V}} a_{\mathcal{V},4} Z_{\mathcal{V},k}$. From the first round transmissions $(X_{1,1}, \ldots, X_{1,4})$, since the coefficient matrix $\mathbf{S}_1$ has rank equal to 4 (by Property 4), the server cannot get any information about W'_k. The second round transmissions remain the same. It can be seen that after two rounds the server can recover $\sum_{k\in[6]} W_{k,j}$ for each $j \in [4]$. In addition, the second round transmissions only contains 4 linear combinations; thus from the second round transmissions, we can at most obtain 4 linear combinations of the input vectors, which are exactly $\sum_{k\in[6]} W_{k,j}$ for each $j \in [4]$. Except these, the server cannot obtain any other information about $W'_1, \ldots, W'_6$. In addition, from $\left(\sum_{k\in[6]} W_{k,j} : j \in [4]\right)$ we only recover $\sum_{k\in[6]} W_{k,i}$ where $i \in [3]$ without the interference from $\sum_{k\in[6]} W_{k,4}$. Hence, the proposed scheme is secure in this case.

Hence, we proved that the proposed scheme is secure when $\mathcal{U}_1 = [6]$. Similarly, the proposed scheme is also secure for other possible $\mathcal{U}_1$. As a result, we achieve the rates $(R_1, R_2) = (1, 1/3)$, coinciding the converse bound in Lemma 1.

V. An Example for Theorem 2

Due to the limitation of pages, we only provide an example to illustrate the main idea of the scheme in Theorem 2.

Example 2 $((\mathsf{K}, \mathsf{U}, \mathsf{S}, \mathsf{T})) = (4, 3, 3, 1))$**.** In this example, we have $\mathsf{S} > \mathsf{K} - \mathsf{U} + 1$ and thus the proposed scheme in Theorem 1 cannot work. Instead, we propose the following secure aggregation scheme. For each $\mathcal{V} \in \binom{[4]}{3}$, we generate $Z_\mathcal{V} = (Z_{\mathcal{V},k} : \mathsf{k} \in \mathcal{V})$ shared by all users in $\mathcal{V}$, where each $Z_{\mathcal{V},k}$ contains L uniformly i.i.d. symbols over $\mathbb{F}_\mathsf{q}$.

First round. We let the users in $[4]$ transmit,

$$\begin{aligned}
X_1 &= W_1 + Z_{\{1,2,3\},1} + Z_{\{1,2,4\},1} + Z_{\{1,3,4\},1},\\
X_2 &= W_2 + Z_{\{1,2,3\},2} + Z_{\{1,2,4\},2} + Z_{\{2,3,4\},2},\\
X_3 &= W_3 + Z_{\{1,2,3\},3} + Z_{\{1,3,4\},3} + Z_{\{2,3,4\},3},\\
X_4 &= W_4 + Z_{\{1,2,4\},4} + Z_{\{1,3,4\},4} + Z_{\{2,3,4\},4},
\end{aligned}$$

respectively. X_k contains L symbols, leading to $R_1 = 1$. Even if the server colludes with one user, the server cannot get any information about W_k from X_k, where user k is not the colluding user. For example, if the server colludes with user 4, W_1 is perfectly protect by $Z_{\{1,2,3\},1}$ and thus secure.

Second round. Recall that the definition of $Z_\mathcal{V}^{\mathcal{U}_1}$ is given in (10). We divide $Z_{\{1,2,3\}}^{\mathcal{U}_1}$ into $\mathsf{U} + \mathsf{S} - \mathsf{K} = 2$ pieces $\{Z_{\{1,2,3\},1}^{\mathcal{U}_1}, Z_{\{1,2,3\},2}^{\mathcal{U}_1}\}$. In order to let the server recover $W_1 + \cdots + W_4$, each user $k \in \mathcal{U}_1$ sends $Y_k^{\mathcal{U}_1}$, where

$$\begin{bmatrix} Y_1^{\mathcal{U}_1} \\ Y_2^{\mathcal{U}_1} \\ Y_3^{\mathcal{U}_1} \\ Y_4^{\mathcal{U}_1} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \\ \mathbf{s}_4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ * & * & * & * & * & * & * & * \end{bmatrix} \begin{bmatrix} Z_{\{1,2,3\},1}^{\mathcal{U}_1} \\ \vdots \\ Z_{\{2,3,4\},1}^{\mathcal{U}_1} \\ Z_{\{1,2,3\},2}^{\mathcal{U}_1} \\ \vdots \\ Z_{\{2,3,4\},2}^{\mathcal{U}_1} \end{bmatrix}.$$

where $[\mathbf{s}_1; \mathbf{s}_2; \mathbf{s}_3; \mathbf{s}_4]$ is any MDS matrix with dimension 4×3, and each '$*$' represents a coefficient to be determined. Note that $Z_{\{1,2,3\}}^{\mathcal{U}_1}$ cannot be encoded by user 4; thus $\mathbf{s}_4 \left[1, 0, *\right]^T = 0$, from which we can fix this '$*$'. Similarly, we can determine every '$*$'.

The decodability is directly from the fact that $[\mathbf{s}_1; \mathbf{s}_2; \mathbf{s}_3; \mathbf{s}_4]$ is an MDS matrix. The security constraint can be proved by the similar method provided in Example 1. Hence, the proposed scheme is decodable and secure, with $R_1 = 1$ and $R_2 = 1/2$.

Acknowledgement: The work of Z. Zhang and K. Wan was partially funded by the National Natural Science Foundation of China NSFC-12141107. The work of M. Ji was supported in part by National Science Foundation (NSF) CAREER Award 2145835. The work of H. Sun was supported in part by NSF Awards 2007108 and 2045656. The work of G. Caire was partially funded by the ERC Advanced Grant N. 789190, CARENET.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, p. 12, 2019.

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60.

[4] H. B. McMahan *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1, 2021.

[5] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.

[6] Y. Zhao and H. Sun, "Information theoretic secure aggregation with user dropouts," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7471–7484, Nov. 2022.

[7] J. So, C. He, C.-S. Yang, S. Li, Q. Yu, R. E. Ali, B. Guler, and S. Avestimehr, "LightSecAgg: a lightweight and versatile design for secure aggregation in federated learning," *arXiv:2109.14236*, Feb. 2022.

[8] K. Wan, H. Sun, M. Ji, and G. Caire, "On the information theoretic secure aggregation with uncoded groupwise keys," *arXiv:2204.11364*, App. 2022.

[9] R. Tandon, Q. Lei, A. G. Dimakis, and N. Karampatziakis, "Gradient coding: Avoiding stragglers in distributed learning," in *International Conference on Machine Learning*. PMLR, 2017, pp. 3368–3376.

[10] M. Ye and E. Abbe, "Communication-computation efficient gradient coding," in *International Conference on Machine Learning*. PMLR, 2018, pp. 5610–5619.

[11] H. Cao, Q. Yan, and X. Tang, "Adaptive gradient coding," *IEEE/ACM Trans. Networking*, vol. 30, no. 2, pp. 717–734, Apr. 2022.