Secure Summation with User Selection and Collusion

Yizhou Zhao
College of Electronic and Information Engineering
Southwest University, Chongqing, China
onezhou@swu.edu.cn

Hua Sun
Department of Electrical Engineering
University of North Texas, Denton, TX USA
hua.sun@unt.edu

2024 IEEE Information Theory Workshop (ITW) | 979-8-3503-4893-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ITW61385.2024.10806928

***Abstract*—The secure summation problem is studied with user selection and collusion, where a server may select any U out of K users and compute the sum of the inputs from the selected users without learning any additional information even if the server colludes with any T out of K users. The optimal communication and randomness rate is characterized when either $U = 2$ or $T = 1$, i.e., to securely compute 1 bit of the selected sum, each user needs to send 1 bit to the server, each user needs to hold a key of $T+1$ bits when $U = 2$ and $U/(U-1)$ bits when $T = 1$, and all users need to hold key variables of $\binom{T+2}{2}$ bits when $U = 2$ and $U/(U-1)+U-1$ bits when $T = 1$.**

I. Introduction

Secure summation [1] arises as a useful Shannon theoretic primitive to study the core security challenges in federated learning [2], [3], i.e., how to enable a server to learn the sum of gradient inputs from K distributed users and nothing more. The focus of this work is on a variant of secure summation that allows user selection, i.e., instead of computing the sum of inputs from all users, the server may select a subset of users and only compute their input sum. When the selected user set can be arbitrary, i.e., ranging from any two users to all users, we show that the minimum key size that each user needs to hold is the harmonic number, $1 + 1/2 + \cdots + 1/(K-1)$ [4].

In this work, we further incorporate the element of user collusion to the secure summation problem with user selection, i.e., we wish to guarantee that the server does not obtain anything beyond the desired sum even if colluding with some users. It turns out that the problem becomes significantly more intricate and we concentrate on a symmetric model where the server may select any U out of the total K users and may collude with any T out of K users, i.e., the cardinality of the selected (colluding) user set is a constant. To securely compute the sum of U inputs, say $W_1+W_2+\cdots+W_U$, a typical protocol is to use zero-sum randomness, i.e., protect each W_k with a key variable S_k (sending $W_k + S_k$) and ensure $S_1 + \cdots + S_U = 0$ (zero-sum property) so that $\sum_{k=1}^{U}(W_k + S_k) = \sum_{k=1}^{U} W_k$ gives us the desired sum. Furthermore, except from the zero-sum property, the key variables S_k should behave generically (e.g., any $U-1$ are independent) to ensure security, i.e., the server cannot infer anything beyond the sum. Along this line, for the considered secure summation problem with user selection and collusion, the technical crux is to ensure that any U users can generate some generic zero-sum randomness keys that are independent of the keys at any T colluding users. We wish to understand the most efficient way of such construction, i.e., the minimum size of the keys at the user side.

The main result of this work includes the exact information theoretic answer for two settings, $U = 2$ or $T = 1$. When $U = 2$ (i.e., each pair of users may be selected), we wish to guarantee that each pair of users may generate a common key (i.e., zero-sum randomness) that is independent of the keys known to any T colluding users. Interpreting in this manner, naturally the code construction from key distribution literature [5] is useful and the converse proof can also be adapted to our secure summation context relying on tools developed in [1], [4]. More broadly and perhaps interestingly, the code construction, which generates generic pairwise overlaps, is identical to that of minimum bandwidth regenerating codes in repairable distributed storage [6]. The detailed result is presented in Theorem 1. When $T = 1$ (i.e., at most 1 user may collude with the server), random linear codes suffice whose dimension design is guided by the converse bounds. The detailed result is presented in Theorem 2.

II. Problem Statement

Consider K users, where User $k \in [1, 2, \cdots, K] \triangleq [K]$ holds an input W_k and a key Z_k. Each independent input W_k is an $L \times 1$ vector with i.i.d. uniform elements from the finite field $\mathbb{F}_q$. Inputs $(W_k)_{k\in[K]}$ are independent of keys $(Z_k)_{k\in[K]}$.

$$H\left((W_k)_{k\in[K]}, (Z_k)_{k\in[K]}\right) = H\left((Z_k)_{k\in[K]}\right) + \sum_{k\in[K]} H(W_k) \quad (= KL \text{ (in } q\text{-ary units)}). \tag{1}$$

Each Z_k is comprised of L_Z symbols from $\mathbb{F}_q$. $(Z_k)_{k\in[K]}$ can be arbitrarily correlated and are a function of a source key variable Z_Σ, which is comprised of L_{Z_Σ} symbols from $\mathbb{F}_q$.

$$H\left((Z_k)_{k\in[K]} \middle| Z_\Sigma\right) = 0. \tag{2}$$

Consider a server who may select an arbitrary set of U users $\mathcal{U}$, where $\mathcal{U} \subset [K]$, $|\mathcal{U}| = U$, and wish to securely compute $\sum_{k\in\mathcal{U}} W_k$. To this end, User $k \in \mathcal{U}$ sends a message $X_k^{\mathcal{U}}$ to the server. The message $X_k^{\mathcal{U}}$ is a function of W_k, Z_k and is comprised of L_X symbols from $\mathbb{F}_q$.

$$H\left(X_k^{\mathcal{U}} | W_k, Z_k\right) = 0, \forall k \in \mathcal{U}. \tag{3}$$

979-8-3503-4893-4/24/$31.00 ©2024 IEEE

From the messages received from the selected users, the server must be able to decode the desired sum $\sum_{k\in\mathcal{U}} W_k$ while nothing more is revealed even if the server may collude any set of at most $T \leq K-U$ users[1] $\mathcal{T}$, where $\mathcal{T} \subset [K], 0 \leq |\mathcal{T}| \leq T$.

$$\text{[Correctness]} \quad H\left(\sum_{k\in\mathcal{U}} W_k \middle| \left(X_k^{\mathcal{U}}\right)_{k\in\mathcal{U}}\right) = 0. \tag{4}$$

[Security] (5)

$$I\left((W_k)_{k\in[K]}; \left(X_k^{\mathcal{U}}\right)_{k\in\mathcal{U}} \middle| \sum_{k\in\mathcal{U}} W_k, (W_k, Z_k)_{k\in\mathcal{T}}\right) = 0.$$

The communication and randomness consumption is measured by the communication rate R, the individual key rate R_Z, and the total key rate R_{Z_Σ}, defined as follows.

$$R \triangleq \frac{L_X}{L}, R_Z \triangleq \frac{L_Z}{L}, \ R_{Z_\Sigma} \triangleq \frac{L_{Z_\Sigma}}{L} \tag{6}$$

which characterizes the normalized number of symbols each message, each key, and source key contains, respectively. A rate tuple (R, R_Z, R_{Z_Σ}) is said to be achievable if there exists a secure summation scheme, for which correctness and security are satisfied, and the communication rate, individual key rate, and total key rate are no greater than R, R_Z, and R_{Z_Σ}, respectively. The closure of the set of all achievable rate tuples is called the optimal rate region, denoted as $\mathcal{R}^*$.

III. Main Result

In this section, we state our main result in the following two theorems, along with essential observations and intuition.

Theorem 1: For K-user secure summation with $U = 2$ selected users and at most $T \leq K-2$ colluding users,

$$\mathcal{R}^* = \left\{(R, R_Z, R_{Z_\Sigma}) : \begin{array}{c} R \geq 1, \\ R_Z \geq T+1, \end{array} R_{Z_\Sigma} \geq \binom{T+2}{2} \right\}.$$

Note that the optimal rate region in Theorem 1 does not depend on K and the rates may simultaneously achieve the minimum, i.e., the rate region is rectangular. To understand the result, first suppose $K = T+2$. For any pair of the $T+2$ users, we assign both users an independent common key symbol so that each user holds $\binom{T+1}{1}$ key symbols and in total we consume $\binom{T+2}{2}$ key symbols. Equipped with such key variables, the code construction is immediate (using zero-sum randomness). The converse essentially shows that such key assignment is necessary (minimum). Proceed now to the $K > T+2$ case where converse continues to hold as more users cannot help and achievability relies on the ingenious construction from key distribution [5] or minimum bandwidth regenerating codes [6] indicating the above $K = T+2$ construction scales to any larger K with exactly the same performance. The detailed proof of Theorem 1 is presented in Section IV.

Theorem 2: For K-user secure summation with U selected users and at most $T = 1 \leq K-U$ colluding user,

$$\mathcal{R}^* = \left\{(R, R_Z, R_{Z_\Sigma}) : \begin{array}{c} R \geq 1, R_Z \geq U/(U-1), \\ R_{Z_\Sigma} \geq U/(U-1) + U - 1 \end{array} \right\}.$$

Due to space limitation, the detailed proof of Theorem 2 is deferred to the full version of this paper and here we give an outline. First, consider the converse. From the total key rate result of secure summation with U users (refer to Theorem 1 of [1]), we have $H(Z_2, \cdots, Z_U|Z_1) \geq (U-1)L$. Further, User 1's key must be correlated with the keys at User 2 to U (in the amount of input size) in order for these U users to be able to securely compute their sum, so $I(Z_1; Z_2, \cdots, Z_U) \geq L$ (the proof is similar to that of (95) in [1]). We now have

$$\begin{aligned} UL &\leq I(Z_1; Z_2, \cdots, Z_U) + H(Z_2, \cdots, Z_U|Z_1) \\ &= H(Z_2, \cdots, Z_U) \end{aligned} \tag{7}$$

so that on average each $H(Z_k) \geq U/(U-1)L$ and the individual key rate bound follows. The total key rate bound is then obtained immediately, i.e., $H(Z_\Sigma) \geq H(Z_1, \cdots, Z_U) = H(Z_1) + H(Z_2, \cdots, Z_U|Z_1) \geq (U/(U-1) + U - 1)L$. The proof of the communicate rate bound $R \geq 1$ is almost identical to that of Theorem 1 in [1]. Second, consider the achievability where we assign the random linear key space dimension according to the converse bounds. Specifically, set $L = U-1$ and let each key lie in an $R_Z L = U$ dimensional generic subspace of an $R_{Z_\Sigma} L = U + (U-1)^2 = U^2 - U + 1$ dimensional ambient space. Any individual key will overlap with the collection of $U-1$ other keys in $U^2 - (U^2 - U + 1) = U - 1 = L$ generic dimensions and the overlapping subspaces will be used as the zero-sum randomness keys to perform secure summation. Security follows from the generic property of the spaces, i.e., the key spaces used by any U users in secure summation with $(U-1)^2$ dimensions are independent of the U dimensional key known to any single colluding user.

IV. Proof of Theorem 1

A. Converse Proof

The proof of $R \geq 1$ is similar to that of Theorem 1 in [1] and is thus omitted. Intuitively, each user needs to send out its input so that $L_X \geq L$, i.e., $R \geq 1$.

We proceed to the key rate bounds. Let us start with some preliminary results. As $U = 2$ and each pair of users may be selected for secure summation, we show that any two keys must share L symbols (captured in a mutual information term, given any colluding user set), in the following lemma.

Lemma 1: For any $i, j \in [K]$, any $\mathcal{T} \subset [K]$, $|\mathcal{T}| \leq T$, $i, j \notin \mathcal{T}$, we have

$$I\left(Z_i; Z_j | (Z_k)_{k\in\mathcal{T}}\right) \geq L. \tag{8}$$

Proof: From the security constraint (5), we have

$$\begin{aligned} 0 &= I\left(W_i, W_j; X_i^{\{i,j\}}, X_j^{\{i,j\}} | W_i + W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) \\ &\geq I\left(W_i; X_i^{\{i,j\}} | W_i + W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) \end{aligned} \tag{9}$$

[1] Without loss of generality, we assume $T \leq K-U$ because otherwise the colluding user set must overlap with the selected user set, which degenerates the security constraint and boils down to the $T = K-U$ setting.

$$\begin{aligned}
&= H\left(W_i|W_i+W_j,(W_k,Z_k)_{k\in\mathcal{T}}\right)\\
&\quad - H\left(W_i|W_i+W_j,(W_k,Z_k)_{k\in\mathcal{T}},X_i^{\{i,j\}}\right) && (10)\\
&\geq L - H\left(W_i|\left(W_k,Z_k\right)_{k\in\mathcal{T}},X_i^{\{i,j\}}\right) && (11)\\
&\Rightarrow\ H\left(W_i|\left(W_k,Z_k\right)_{k\in\mathcal{T}},X_i^{\{i,j\}}\right)\geq L && (12)
\end{aligned}$$

where to obtain the first term of (11), we use the fact that each input of L uniform symbols is independent of other inputs and keys (see (1)). Then

$$\begin{aligned}
&I\left(Z_i;Z_j|\left(Z_k\right)_{k\in\mathcal{T}}\right)\\
&= I\left(W_i,Z_i;W_j,Z_j|\left(W_k,Z_k\right)_{k\in\mathcal{T}}\right) && (13)\\
&\overset{(3)}{\geq} I\left(W_i,X_i^{\{i,j\}};W_j,X_j^{\{i,j\}}|\left(W_k,Z_k\right)_{k\in\mathcal{T}}\right) && (14)\\
&\geq I\left(W_i;W_j,X_j^{\{i,j\}}|\left(W_k,Z_k\right)_{k\in\mathcal{T}},X_i^{\{i,j\}}\right) && (15)\\
&= H\left(W_i|\left(W_k,Z_k\right)_{k\in\mathcal{T}},X_i^{\{i,j\}}\right)\\
&\quad -\ H\left(W_i|\left(W_k,Z_k\right)_{k\in\mathcal{T}},X_i^{\{i,j\}},W_j,X_j^{\{i,j\}}\right) && (16)\\
&\overset{(12)}{\geq}\ L && (17)
\end{aligned}$$

where (13) follows from the independence of the inputs and keys (refer to (1)). In (14), we use the fact that $X_k^{\mathcal{U}}$ is a function of W_k, Z_k. The second term of (16) is 0 because by the correctness constraint (4), $W_i + W_j$ can be obtained from $X_i^{\{i,j\}}, X_j^{\{i,j\}}$ and then combining with W_j (available in the conditional terms), we can recover W_i. ■

To prepare for the converse proof, we obtain a consequence of Lemma 1, stated in the following lemma.

Lemma 2: For any $i \in [K]$, any $\mathcal{T} \subset [K]$, $|\mathcal{T}| \leq T$, $i \notin \mathcal{T}$, we have

$$H\left(Z_i|(Z_k)_{k\in\mathcal{T}}\right) \geq (T-|\mathcal{T}|+1)L. \tag{18}$$

Proof: Without loss of generality, set $i = 1$ and $\mathcal{T} = \{K-|\mathcal{T}|+1,\cdots,K-1,K\}$.

$$\begin{aligned}
&H\left(Z_1|(Z_k)_{k\in\mathcal{T}}\right)\\
\geq\ & I\left(Z_1;Z_2,\cdots,Z_{K-|\mathcal{T}|}|(Z_k)_{k\in\mathcal{T}}\right) && (19)\\
=\ & I\left(Z_1;Z_2|(Z_k)_{k\in\mathcal{T}}\right)+I\left(Z_1;Z_3|(Z_k)_{k\in\mathcal{T}},Z_2\right)+\cdots\\
& + I\left(Z_1;Z_{K-|\mathcal{T}|}|(Z_k)_{k\in\mathcal{T}},Z_2,\cdots,Z_{K-|\mathcal{T}|-1}\right) && (20)\\
\overset{(8)}{\geq}\ & (T-|\mathcal{T}|+1)L && (21)
\end{aligned}$$

where to obtain (21), the first $T-|\mathcal{T}|+1$ terms of (20) are bounded by Lemma 1 (note that the choice of $\mathcal{T}$ may need to vary when bounding different terms), and remaining terms are non-negative. ■

We are now ready to prove the key rate converse. First, we apply Lemma 2 through setting $\mathcal{T} = \emptyset$ to show $R_Z \geq T+1$.

$$L_Z \geq H(Z_k) \overset{(18)}{\geq} (T+1)L\ \Rightarrow\ R_Z = L_Z/L \geq T+1. \tag{22}$$

Second, consider the total key rate bound $R_{Z_\Sigma} \geq \binom{T+2}{2}$.

$$\begin{aligned}
L_{Z_\Sigma}\ &\geq\ H(Z_\Sigma)\geq H\left(Z_1,Z_2,\cdots,Z_{T+2}\right) && (23)\\
&\geq\ H\left(Z_1\right)+H\left(Z_2|Z_1\right)+\cdots\\
&\quad\ + H\left(Z_{T+2}|Z_1,\cdots,Z_{T+1}\right) && (24)\\
&\overset{(18)}{\geq}\ (T+1)L+TL+\cdots+L+0 && (25)\\
&=\ \binom{T+2}{2}L && (26)\\
\Rightarrow\quad R_{Z_\Sigma}\ &=\ L_{Z_\Sigma}/L\geq\binom{T+2}{2}. && (27)
\end{aligned}$$

B. Achievability Proof of Example $K = 5, U = 2, T = 2$

We first present the achievable scheme for an example to illustrate the idea in a simpler setting. Consider $K = 5$ users, where any $U = 2$ users may be selected and any $T = 2$ users may collude with the server. We show that the rates $R = 1, R_Z = 3, R_{Z_\Sigma} = 6$ are achievable[2].

Suppose $L = B$, i.e., each input W_k contains B symbols from $\mathbb{F}_q$. Equivalently, we may view W_k as one symbol from the extension field $\mathbb{F}_{q^B}$. Our scheme will operate over the extension field $\mathbb{F}_{q^B}$ and it is useful for now to think of B as a large integer so that we are working in a sufficiently large field (an exact choice of B will be given in the general proof). Consider a 6×1 vector over $\mathbb{F}_{q^B}$, $S = [S_1, S_2, \cdots, S_6]^\top$ where S_i are i.i.d. uniform and define a symmetric matrix

$$\mathbf{S} \triangleq \begin{bmatrix} S_1 & S_2 & S_3\\ S_2 & S_4 & S_5\\ S_3 & S_5 & S_6 \end{bmatrix}. \tag{28}$$

Set $A_k = [a_k^1, a_k^2, a_k^3]^\top \in \mathbb{F}_{q^B}^{3\times 1}, k \in [5]$. The vectors A_k are chosen before the communication protocol starts and are known to all users as global codebook knowledge. We will show that there exists a choice of A_k that will produce a correct and secure achievable scheme. Each individual key Z_k is set as

$$Z_k \quad = \quad \mathbf{S}A_k = \begin{bmatrix} S_1 & S_2 & S_3\\ S_2 & S_4 & S_5\\ S_3 & S_5 & S_6 \end{bmatrix}\begin{bmatrix} a_k^1\\ a_k^2\\ a_k^3 \end{bmatrix}, \tag{29}$$

or equivalently,

$$\begin{aligned}
Z_k \quad &= \quad \begin{bmatrix} a_k^1 & a_k^2 & a_k^3 & 0 & 0 & 0\\ 0 & a_k^1 & 0 & a_k^2 & a_k^3 & 0\\ 0 & 0 & a_k^1 & 0 & a_k^2 & a_k^3 \end{bmatrix}\begin{bmatrix} S_1\\ S_2\\ S_3\\ S_4\\ S_5\\ S_6 \end{bmatrix} && (30)\\
&\triangleq \quad \mathbf{A}_k S. && (31)
\end{aligned}$$

Note that each Z_k contains $L_Z = 3B$ symbols from $\mathbb{F}_q$, so $R_Z = L_Z/L = 3$; the source key variable S contains $L_{Z_\Sigma} = 6B$ symbols from $\mathbb{F}_q$, so $R_{Z_\Sigma} = L_{Z_\Sigma}/L = 6$.

[2] We present the achievable scheme using the product matrix framework from regenerating codes literature [6] but note that the polynomial based approach from key distribution literature will work equally well [5].

For any $U = 2$ selected users $i, j \in [5], i < j$, the messages are set as

$$\begin{aligned} X_i^{\{i,j\}} &= W_i + A_j^\top Z_i = W_i + A_j^\top \mathbf{S} A_i, \\ X_j^{\{i,j\}} &= W_j - A_i^\top Z_j = W_j - A_i^\top \mathbf{S} A_j \end{aligned} \tag{32}$$

where $A_j^\top \mathbf{S} A_i = A_i^\top \mathbf{S} A_j$ as $\mathbf{S}$ is symmetric, i.e., $\mathbf{S} = \mathbf{S}^\top$. Correctness is guaranteed as $W_i + W_j = X_i^{\{i,j\}} + X_j^{\{i,j\}}$. Note that each message contains $L_X = B$ symbols from $\mathbb{F}_q$, so $R = L_X / L = 1$.

Finally, we are left with the security proof, which is the most technical part. We will see that the security constraint (5) will boil down to requiring that a number of matrices in variables A_k have full rank. Therefore, if the elements of A_k are chosen in a generic manner, the full rank condition will be satisfied with non-zero probability (if B is chosen sufficiently large) thus guaranteeing the existence of a feasible solution.

To see the full rank condition more concretely, let us consider an example of the security constraint (5) when $\mathcal{U} = \{1, 2\}$ and $\mathcal{T} = \{3, 4\}$,

$$\begin{aligned} & I\left(W_1, W_2; X_1^{\{1,2\}}, X_2^{\{1,2\}} | W_1 + W_2, W_3, Z_3, W_4, Z_4\right) \\ &= I\left(W_1, W_2; W_1 + A_2^\top \mathbf{S} A_1 | W_1 + W_2, W_3, Z_3, W_4, Z_4\right) && (33) \\ &= H\left(W_1 + A_2^\top \mathbf{S} A_1 | W_1 + W_2, W_3, Z_3, W_4, Z_4\right) \\ &\quad - H\left(A_2^\top \mathbf{S} A_1 | W_1, W_2, W_3, Z_3, W_4, Z_4\right) && (34) \\ &\leq B - H\left(A_2^\top \mathbf{S} A_1 | \mathbf{S} A_3, \mathbf{S} A_4\right) && (35) \\ &= B - H\left(A_2^\top \mathbf{S} A_1, \mathbf{S} A_3, \mathbf{S} A_4\right) + H\left(\mathbf{S} A_3, \mathbf{S} A_4\right) && (36) \\ &= B - 6B + 5B && (37) \\ &= 0 && (38) \end{aligned}$$

where in (33), $X_2^{\{1,2\}}$ is removed since $X_2^{\{1,2\}} = W_1 + W_2 - X_1^{\{1,2\}}$, and we plug in $X_1^{\{1,2\}}$ (see (32)). In (35), the first term follows from the fact that $W_1 + A_2^\top \mathbf{S} A_1$ contains B symbols from $\mathbb{F}_q$, whose entropy is at most B in q-ary units; the second term follows from the independence of the keys and inputs (refer to (1)). To obtain (37), we are left to show that there exists a choice of A_k so that $H\left(A_2^\top \mathbf{S} A_1, \mathbf{S} A_3, \mathbf{S} A_4\right) = 6B$ and $H\left(\mathbf{S} A_3, \mathbf{S} A_4\right) = 5B$. We will see that when A_k is sufficiently generic, e.g., i.i.d. uniform over a large field, then $A_2^\top \mathbf{S} A_1$ will likely be independent of $\mathbf{S} A_3, \mathbf{S} A_4$ and the desired entropy equality holds.

First, consider $H\left(A_2^\top \mathbf{S} A_1, \mathbf{S} A_3, \mathbf{S} A_4\right)$.

$$H\left(A_2^\top \mathbf{S} A_1, \mathbf{S} A_3, \mathbf{S} A_4\right) = H\left(\left[\begin{array}{c} \mathbf{A}_3 \\ \mathbf{A}_4 \\ \mathbf{A}_{1\cap 2} \end{array}\right] S\right) \tag{39}$$

where $\mathbf{A}_{1\cap 2}$ denotes the 1 dimensional overlap of Z_1 and Z_2 and is defined as

$$\mathbf{A}_{1\cap 2} S \triangleq A_2^\top \mathbf{S} A_1. \tag{40}$$

In this case, it can be shown that

$$\begin{aligned} \mathbf{A}_{1\cap 2} \quad = \quad & [a_1^1 a_2^1, a_1^1 a_2^2 + a_1^2 a_2^1, a_1^3 a_2^1 + a_1^1 a_2^3, \\ & a_1^2 a_2^2, a_1^3 a_2^2 + a_1^2 a_2^3, a_1^3 a_2^3]. \end{aligned} \tag{41}$$

To ensure that $H\left(A_2^\top \mathbf{S} A_1, \mathbf{S} A_3, \mathbf{S} A_4\right) = 6B$, we need to guarantee matrix $\mathbf{A} = [\mathbf{A}_3; \mathbf{A}_4; \mathbf{A}_{1\cap 2}]$ has full rank of 6 over $\mathbb{F}_{q^B}$ - this is the full rank condition we mentioned earlier. To see that this full rank condition can be satisfied, view the determinant of a square sub-matrix of $\mathbf{A}$ as a polynomial in variables $a_k^i, k \in [5], i \in [3]$. As long as the determinant polynomial is not identically zero, then by Schwartz–Zippel lemma, when the field size q^B is sufficiently large, there must exist a realization of the above matrix that has full rank. To show the determinant is not the zero polynomial, it suffices to give a realization such that $\mathbf{A}$ contains a full rank 6×6 sub-matrix. Set $(a_3^1, a_3^2, a_3^3) = (1, 0, 0)$, $(a_4^1, a_4^2, a_4^3) = (0, 1, 0)$, and $a_1^3 a_2^3 = 1$ so that

$$\mathbf{A} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \multicolumn{5}{c}{\cdots} & 1 \end{array}\right] \tag{42}$$

where the elements in dots can be chosen arbitrarily and note that the 6×6 submatrix $\mathbf{A}(1, 2, 3, 5, 6, 7; :)$ has full rank (i.e., all rows except the 4-th, which is lower triangular with diagonal entries being 1).

Second, consider $H\left(\mathbf{S} A_3, \mathbf{S} A_4\right) = H([\mathbf{A}_3; \mathbf{A}_4] S)$, which can be considered similarly as above. In fact, the exactly same choice of a_k^i will ensure that rank$[\mathbf{A}_3; \mathbf{A}_4] = 5$ over $\mathbb{F}_{q^B}$ as $[\mathbf{A}_3; \mathbf{A}_4]$ is a sub-matrix of $\mathbf{A}$ (i.e., the first 6 rows), so that $H\left(\mathbf{S} A_3, \mathbf{S} A_4\right) = 5B$.

We have finished considering one choice of $\mathcal{U}$ and $\mathcal{T}$. In general we need to guarantee security for all choices of $\mathcal{U}$ and $\mathcal{T}$, each of which will induce a full rank condition. As long as we can guarantee that the determinant polynomial in each case is not the zero polynomial, then by Schwartz–Zippel lemma, when the field size q^B is larger than the degree of the product of all determinant polynomials, there exists a choice of a_k^i so that all full rank conditions are satisfied, i.e., security is guaranteed. We proceed next to the general proof.

C. General Achievability Proof

The general achievability proof is a generalization of that of the above example and here we highlight the differences.

Suppose $L = B$, where[3] $q^B > (1 + 2T^2) K^2 2^K$ and we operate over field $\mathbb{F}_{q^B}$. Consider $S = [S_1, S_2, \cdots, S_{\binom{T+2}{2}}]^\top$ whose elements are i.i.d. uniform over $\mathbb{F}_{q^B}$ and define $(T + 1) \times (T + 1)$ symmetric matrix

$$\mathbf{S} \triangleq \left[\begin{array}{cccc} S_1 & S_2 & \cdots & S_{T+1} \\ S_2 & S_{T+2} & & S_{2T+1} \\ \vdots & & \ddots & \vdots \\ S_{T+1} & S_{2T+1} & \cdots & S_{\binom{T+2}{2}} \end{array}\right]. \tag{43}$$

[3] The block size B is exceedingly large and we view it as a Shannon style existence proof (B is not optimized).

Set $A_k = [a_k^1, a_k^2, \cdots, a_k^{T+1}]^\top \in \mathbb{F}_{q^B}^{(T+1)\times 1}, k \in [K]$ and each individual key as

$$Z_k = \mathbf{S}A_k \triangleq \mathbf{A}_k S. \tag{44}$$

For any $U = 2$ selected users $i, j \in [K], i < j$, the messages are set as

$$\begin{aligned} X_i^{\{i,j\}} &= W_i + A_j^\top Z_i = W_i + A_j^\top \mathbf{S} A_i, \\ X_j^{\{i,j\}} &= W_j - A_i^\top Z_j = W_j - A_i^\top \mathbf{S} A_j \end{aligned} \tag{45}$$

so $W_i + W_j = X_i^{\{i,j\}} + X_j^{\{i,j\}}$ (note that $A_j^\top \mathbf{S} A_i = A_i^\top \mathbf{S} A_j$ due to the symmetry of $\mathbf{S}$) and correctness is guaranteed.

Next, we proceed to the security proof (by showing existence of A_k) and capture the crucial property in the following lemma. Define $\mathbf{A}_{i\cap j} S \triangleq A_i^\top \mathbf{S} A_j$.

Lemma 3: When $q^B > (1+2T^2)K^2 2^K$, there exists a choice of $A_k, k \in [K]$ such that $\forall \mathcal{T}, |\mathcal{T}| \leq T, \forall i, j \notin \mathcal{T}$,

$$H\left(\mathbf{A}_{i\cap j} S, (Z_k)_{k\in\mathcal{T}}\right) \;\geq\; \left(1 + \frac{1}{2}\Big(2T + 3 - |\mathcal{T}|\Big)|\mathcal{T}|\right) B, \tag{46}$$

$$H\left((Z_k)_{k\in\mathcal{T}}\right) \;\leq\; \left(\frac{1}{2}\Big(2T + 3 - |\mathcal{T}|\Big)|\mathcal{T}|\right) B, \tag{47}$$

$$H\left(\mathbf{A}_{i\cap j} S | \, (Z_k)_{k\in\mathcal{T}}\right) \;\geq\; B. \tag{48}$$

Remark: The three inequalities in Lemma 3 can be shown to be equalities, but the inequality form suffices for our proof.

Proof: First, consider (46). Denote $\mathcal{T} = \{t_1, t_2, \cdots, t_{|\mathcal{T}|}\}$.

$$H\left(\mathbf{A}_{i\cap j} S, (Z_k)_{k\in\mathcal{T}}\right) = H\left(\begin{bmatrix} \mathbf{A}_{t_1} \\ \mathbf{A}_{t_2} \\ \vdots \\ \mathbf{A}_{t_{|\mathcal{T}|}} \\ \mathbf{A}_{i\cap j} \end{bmatrix} S\right) \triangleq H(\mathbf{A}S).$$

It suffices to show that we may set A_k so that the rank of $\mathbf{A}$ is at least $1 + \frac{1}{2}(2T + 3 - |\mathcal{T}|)|\mathcal{T}|, \forall \mathcal{T}, |\mathcal{T}| \leq T, \forall i, j \notin \mathcal{T}$ over $\mathbb{F}_{q^B}$. Similar to the proof of the previous example, we show that for each $\mathbf{A}$, the determinant polynomial of a square submatrix is not identically zero. To this end, set

$$a_{t_m}^n = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}, \forall m \in [|\mathcal{T}|], \tag{49}$$

$$a_i^{T+1} = a_j^{T+1} = 1, \tag{50}$$

and similar to (42), $\mathbf{A}$ contains a square submatrix which is lower triangular and all diagonal entries are 1; the dimension of the square matrix is $1 + (T+1) + T + (T-1) + \cdots + (T + 2 - |\mathcal{T}|) = 1 + \frac{1}{2}(2T + 3 - |\mathcal{T}|)|\mathcal{T}|$. Note that the dimension is no greater than $1 + 2T^2$ and there are at most $K^2 2^K$ such polynomials, as the number of choices of i, j is no greater than K^2 and the number of choices of $\mathcal{T}$ is no greater than 2^K. Consider now the product of all such non-zero determinant polynomials, whose degree is at most $(1 + 2T^2)K^2 2^K < q^B$. By Schwartz–Zippel lemma, there exists a choice of A_k for which (46) holds. Fix now the choice of A_k and next we show that for this choice of A_k, (47) and (48) are also valid.

Second, consider (47). The proof is based on mathematical induction on $|\mathcal{T}|$.

Base case: when $|\mathcal{T}| = 1$, $H(Z_k) \leq (T+1)B$ as from (44), Z_k contains $T + 1$ symbols from $\mathbb{F}_{q^B}$.

Induction step: Suppose (47) holds $\forall \mathcal{T}, |\mathcal{T}| \leq m < T$ and we need to show that (47) also holds $\forall \mathcal{T}', |\mathcal{T}'| = m + 1 \leq T$. Denote $\mathcal{T}' = \{t_1, t_2, \cdots, t_{m+1}\}$.

$$\begin{aligned}
& H\left((Z_k)_{k\in\mathcal{T}'}\right) = H(Z_{t_1}, \cdots, Z_{t_{m+1}}) && (51) \\
&= H\left(Z_{t_1}, \cdots, Z_{t_m}\right) + H\left(Z_{t_{m+1}} | Z_{t_1}, \cdots, Z_{t_m}\right) && (52) \\
&= H\left(Z_{t_1}, \cdots, Z_{t_m}\right) + H\left(Z_{t_{m+1}}\right) \\
&\quad - \sum_{i\in[m]} I\left(Z_{t_{m+1}}; Z_{t_i} | (Z_{t_j})_{j\in[i-1]}\right) && (53) \\
&\leq H\left(Z_{t_1}, \cdots, Z_{t_m}\right) + H\left(Z_{t_{m+1}}\right) \\
&\quad - \sum_{i\in[m]} H\left(\mathbf{A}_{t_{m+1}\cap t_i} S | (Z_{t_j})_{j\in[i-1]}\right) && (54) \\
&= H\left(Z_{t_1}, \cdots, Z_{t_m}\right) + H\left(Z_{t_{m+1}}\right) \\
&\quad - \sum_{i\in[m]} \Bigg(H\left(\mathbf{A}_{t_{m+1}\cap t_i} S, (Z_{t_j})_{j\in[i-1]}\right) \\
&\qquad\qquad - H\left((Z_{t_j})_{j\in[i-1]}\right) \Bigg) && (55) \\
&\overset{(46)}{\leq} \left(\frac{1}{2}\Big(2T + 3 - m\Big)m + (T+1) - m\right) B && (56) \\
&= \left(\frac{1}{2}\Big(2T + 3 - (m+1)\Big)(m+1)\right) B && (57)
\end{aligned}$$

where (54) follows from our key assignment (44) so that $Z_{t_{m+1}}$ and Z_{t_i} share $\mathbf{A}_{t_{m+1}\cap t_i} S$. To obtain (56), we plug in the induction assumption on $H\left((Z_k)_{k\in\mathcal{T}}\right)$ for all $\mathcal{T}$ such that $|\mathcal{T}| \leq m$ and (46). The proof of (47) is complete.

Finally, (48) is a direct consequence of (46) and (47). ■

We are now ready to verify the security constraint (5).

$$\begin{aligned}
& I\left(W_i, W_j; X_i^{\{i,j\}}, X_j^{\{i,j\}} | W_i + W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) \\
&= H\left(X_i^{\{i,j\}}, X_j^{\{i,j\}} | W_i + W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) \\
&\quad - H\left(X_i^{\{i,j\}}, X_j^{\{i,j\}} | W_i, W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) && (58) \\
&\overset{(45),(4)}{=} H\left(W_i + A_j^\top \mathbf{S} A_i | W_i + W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) \\
&\quad - H\left(A_j^\top \mathbf{S} A_i | W_i, W_j, (W_k, Z_k)_{k\in\mathcal{T}}\right) && (59) \\
&\overset{(1)}{\leq} B - H\left(\mathbf{A}_{i\cap j} S | \, (Z_k)_{k\in\mathcal{T}}\right) && (60) \\
&\overset{(48)}{\leq} B - B = 0. && (61)
\end{aligned}$$

Finally, we have $R = L_X/L = 1$, $R_Z = L_Z/L = T + 1$, and $R_{Z_\Sigma} = L_{Z_\Sigma}/L = \binom{T+2}{2}$, as desired.

Acknowledgment

The work of Yizhou Zhao is supported by “Fundamental Research Funds for the Central Universities” under Project No. SWU-KQ24040. The work of Hua Sun is supported in part by NSF under Grant CCF-2007108, Grant CCF-2045656 and Grant CCF-2312228.

References

[1] Y. Zhao and H. Sun, "Secure Summation: Capacity Region, Groupwise Key, and Feasibility," *IEEE Transactions on Information Theory*, vol. 70, no. 2, pp. 1376–1387, 2024.

[2] 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 *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.

[3] Y. Zhao and H. Sun, "Information Theoretic Secure Aggregation With User Dropouts," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7471–7484, 2022.

[4] ——, "MDS Variable Generation and Secure Summation with User Selection," *arXiv preprint arXiv:2211.01220*, 2022.

[5] C. Blundo, A. De Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, "Perfectly-Secure Key Distribution for Dynamic Conferences," in *Annual International Cryptology Conference*. Springer, 1992, pp. 471–486.

[6] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the msr and mbr points via a product-matrix construction," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 5227–5239, 2011.