Secure Summation: Capacity Region, Groupwise Key, and Feasibility

Yizhou Zhao, *Member, IEEE*, and Hua Sun, *Member, IEEE*

***Abstract*— The secure summation problem is considered, where K users, each holds an input, wish to compute the sum of their inputs at a server securely, i.e., without revealing any information beyond the sum even if the server may collude with any set of up to T users. First, we prove a folklore result for secure summation - to compute 1 bit of the sum securely, each user needs to send at least 1 bit to the server, each user needs to hold a key of at least 1 bit, and all users need to hold collectively some key variables of at least $K-1$ bits. Next, we allow any arbitrary group of users to share an independent key and any arbitrary group of users to collude with the server. For such a general groupwise key and colluding user setting, we show that secure summation is feasible if and only if the hypergraph, where each node is a user and each edge is a group of users sharing the same key, is connected after removing the nodes corresponding to any colluding set of users and their incident edges. Finally, we focus on the symmetric groupwise key setting, where every group of G users share an independent key. We show that for symmetric groupwise keys with group size G, if $G=1$ or $G>K-T$, the secure summation problem is not feasible; else $1<G\leq K-T$, to compute 1 bit of the sum securely, each user needs to send at least 1 bit to the server and the size of each groupwise key is at least $(K-T-1)/\binom{K-T}{G}$ bits.**

***Index Terms*— Secure aggregation, information theoretic security, capacity region.**

I. Introduction

THE advent of the modern information age is enabled by pervasive networked communication and computation devices, which accelerate data exchange at an unprecedented pace and bring security concerns to the forefront. The need to securely perform distributed computation tasks has thus increased tremendously. This work is particularly motivated by the secure aggregation problem [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], which arises recently in federated learning and the core is to securely compute the sum of the inputs available at a number of users without revealing any additional information to a server. While secure aggregation is usually involved with more practical elements that are crucial for machine learning applications, such as user dropouts, peer-to-peer communication among the users etc., in this work we focus on an elemental information theoretic model that is possibly the simplest while capturing the core of secure sum computation (referred to as secure summation), and wish to understand its fundamental limits on communication and randomness cost.

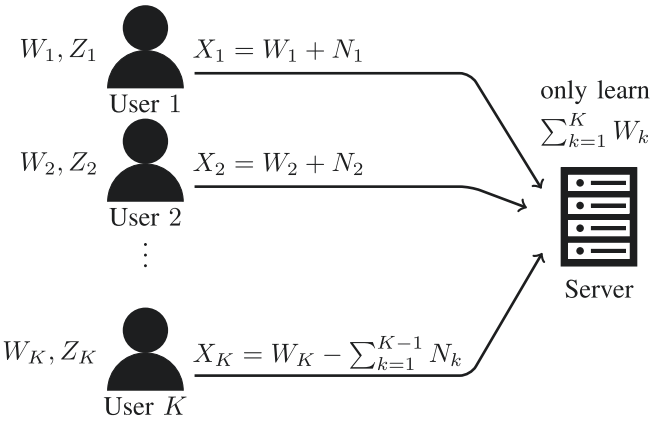

Fig. 1. The secure summation problem and an optimal protocol.

In the secure summation problem (see Fig. 1), we have K users, each holds an independent input $W_k, k \in \{1, 2, \cdots, K\}$ over some finite field and a key variable Z_k that is independent of W_k and serves to ensure security. Each user is connected to the server through an orthogonal noiseless link, over which User k can send a message X_k. After receiving one message from each user (i.e., from $X_1, \cdots, X_K$), the server must be able to compute the sum of all W_k but learn no extra information in the information theoretic sense even if the server may collude with at most[1] T arbitrary users.

A simple protocol (see e.g., [11], [12]) is depicted in Fig. 1, where the key variables are assigned such that their sum is zero, i.e., $Z_k = N_k, k \in \{1, \cdots, K-1\}, Z_K = -\sum_{k=1}^{K-1} Z_k$ and N_k are i.i.d. and uniform over the same field as the inputs. Interestingly, this protocol turns out to be information theoretically optimal in terms of both the communication cost and the key size - in order to securely compute 1 symbol of the sum $W_1 + \cdots + W_K$, each user must communicate at least

Manuscript received 24 October 2022; revised 24 July 2023; accepted 29 November 2023. Date of publication 13 December 2023; date of current version 22 January 2024. This work was supported in part by NSF under Grant CCF-2007108, Grant CCF-2045656, and Grant CCF-2312228. *(Corresponding author: Yizhou Zhao.)*

Yizhou Zhao is with the College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China (e-mail: onezhou@swu.edu.cn).

Hua Sun is with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: hua.sun@unt.edu).

Communicated by R. Tandon, Associate Editor for Security and Privacy.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3342571.

Digital Object Identifier 10.1109/TIT.2023.3342571

[1]Without loss of generality, we assume $0 \leq T \leq K-2$ because when the server colludes with $K-1$ or K users, the server can learn all inputs W_k such that there is nothing to hide. As a result, when $T=K-1$ or K, the results are the same as those when $T=K-2$.

1 symbol of X_k to the server, each user must hold a key Z_k that is at least 1 symbol, and all users must collectively hold key variables whose joint entropy $H(Z_1, \cdots, Z_K)$ is at least $K-1$ symbols. Note that the optimal communication and key rates do not depend on the maximum number of colluding users, T. While we expect this result to be known and indeed the proof follows relatively straightforwardly from existing[2] work (e.g., [5], [13]), we do not find this result explicitly published anywhere and therefore give a full proof in this work (refer to Theorem 1), which also turns out to be useful in our next set of results.

In the basic model of secure summation, the keys Z_k are allowed to be arbitrarily correlated and thus typically are assigned by a trusted third-party or generated through interactive communication among the users, which might be occasionally restrictive in practice. Observing that generating a key securely among a group of users is a well studied primitive [14], [15], [16], [17], [18], we proceed to consider the groupwise key model, where any group of users from the set $\mathcal{G}$ may share an independent key $S_{\mathcal{G}}$. Furthermore, we allow any group of users to be a colluding user set. We wish to understand for which groupwise key and colluding user pattern, the secure summation problem is feasible. The necessary and sufficient condition is obtained in Theorem 2, which is stated in terms of a hypergraph representation of the key pattern, explained now. For example, suppose we have $K = 4$ users and 3 groupwise keys $S_{\{1,2,4\}}$ (i.e., a key shared by User 1, 2, 4), $S_{\{2,3\}}, S_{\{3,4\}}$. Representing this groupwise key setting with a hypergraph (see Fig. 2(a)), we have 4 nodes v_1, v_2, v_3, v_4, where v_k corresponds to User k, and 3 hyperedges e_1, e_2, e_3, each corresponds to a set of users sharing a same key (e.g., e_1 corresponds to $S_{\{1,2,4\}}$ such that e_1 is incident with v_1, v_2, v_4).

Equipped with the hypergraph key representation, we are ready to state the result. We show that the secure summation problem is feasible if and only if after removing the node set that corresponds to any colluding set of users and their incident edges, the remaining hypergraph is connected (refer to Theorem 2). Consider again the example in Fig. 2(a). Suppose User 4 might collude with the server. To see if this secure summation problem is feasible, we remove v_4 and its incident edges from the hypergraph and obtain Fig. 2(b), which is not connected as v_1 is isolated, so we conclude that secure summation is not feasible. The intuition is that there are too few keys that are hidden from the colluding user set so that it is not possible to fully protect the desired sum. As another example, suppose User 3 might be colluding, removing which we have a connected subgraph Fig. 2(c) so that secure summation is feasible. Here we have sufficient keys to ensure both security and decodability of the sum. When there are multiple possible colluding user sets, for a secure summation problem to be feasible, we need to ensure the connectivity of the hypergraph after removing each colluding user set.

[2]Note that the model of [5] assumes some user must drop so that the communication model contains two rounds and it does not allow us to set user dropouts to null to reduce to the secure summation problem studied in this work.

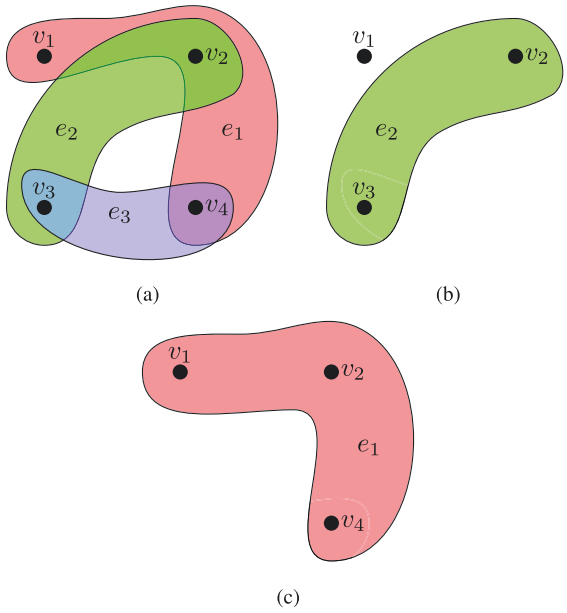

Fig. 2. (a) The hypergraph representation of 3 groupwise keys $S_{\{1,2,4\}}$, $S_{\{2,3\}}$, $S_{\{3,4\}}$. (b) The remaining hypergraph after removing v_4. (c) The remaining hypergraph after removing v_3.

Finally, we consider the symmetric groupwise key setting, where every G users share an independent key and all such keys have the same size. For example, suppose $K = 3, G = 2$. Then we have 3 groupwise keys $S_{\{1,2\}}, S_{\{1,3\}}, S_{\{2,3\}}$ and the key variable at User 1 is $Z_1 = (S_{\{1,2\}}, S_{\{1,3\}})$.

For the secure summation problem with symmetric groupwise keys, we fully characterize the optimal communication and key rates in Theorem 3. We have two regimes - if $G = 1$ or $G > K-T$, the secure summation problem is not information theoretically feasible, intuitively because too many keys are known to a colluding user set such that leakage cannot be avoided; else $1 < G \leq K-T$, to securely compute 1 symbol of the sum, the optimal communication rate remains the same as the arbitrarily coded key case, i.e., groupwise keys do not hurt and each user must communicate at least 1 symbol to the server, and the minimum size of the symmetric groupwise key $S_{\mathcal{G}}$ is $(K-T-1)/\binom{K-T}{G}$ symbols. The achievable scheme is based on a randomized key construction and the converse builds upon that of the arbitrarily coded key setting considered in Theorem 1 and incorporates the groupwise key constraint (refer to Theorem 3).

Notation: For positive integers K_1, K_2, $K_1 < K_2$, we use the notation $[K_1 : K_2] \triangleq \{K_1, K_1+1, \cdots, K_2\}$ and $[1 : K_2]$ is abbreviated as $[K_2]$. The notation $|\mathcal{A}|$ is used to denote the cardinality of a set $\mathcal{A}$. For two sets $\mathcal{A}$ and $\mathcal{B}$, we use $\mathcal{A}\backslash\mathcal{B}$ to denote the set of elements that belong to $\mathcal{A}$ but not $\mathcal{B}$. The notation $\binom{\mathcal{A}}{G}$ is used to denote all subsets of $\mathcal{A}$ with cardinality G, i.e., $\binom{\mathcal{A}}{G} \triangleq \{\mathcal{G} : \mathcal{G} \subset \mathcal{A}, |\mathcal{G}| = G\}$.

II. Problem Statement

The secure summation problem involves one server and K users, where $K \geq 2$ and User $k \in [K]$ holds an input vector W_k and a key variable Z_k. The input vectors $(W_k)_{k\in[K]}$ are

independent. Each W_k is an $L \times 1$ column vector and the L elements are i.i.d. uniform symbols from the finite field $\mathbb{F}_q$. $(W_k)_{k\in[K]}$ is independent of $(Z_k)_{k\in[K]}$.

$$H\left((W_k, Z_k)_{k\in[K]}\right) = \sum_{k\in[K]} H\left(W_k\right) + H\left((Z_k)_{k\in[K]}\right), \quad (1)$$

$$H(W_k) = L \text{ (in } q\text{-ary units)}, \ \forall k \in [K]. \quad (2)$$

Each Z_k is comprised of L_Z symbols from $\mathbb{F}_q$. The key variables can be arbitrarily correlated and are a function of a source key variable Z_Σ, which is comprised of L_{Z_Σ} symbols from $\mathbb{F}_q$.

$$H\left((Z_k)_{k\in[K]}\middle| Z_\Sigma\right) = 0. \quad (3)$$

The communication protocol includes one message from each user to the server. Specifically, User k sends a message X_k, $k \in [K]$ to the server. The message X_k is a function of W_k, Z_k and consists of L_X symbols from $\mathbb{F}_q$.

$$H\left(X_k|W_k, Z_k\right) = 0, \forall k \in [K]. \quad (4)$$

From all messages, the server must be able to recover the desired sum $\sum_{k\in[K]} W_k$ with no error.

$$\text{[Correctness]} \quad H\left(\sum_{k\in[K]} W_k \middle| (X_k)_{k\in[K]}\right) = 0. \quad (5)$$

We impose that security must be guaranteed even if the server may collude with any set of at most T users, where $0 \le T \le K-2$. Specifically, security refers to the constraint that the server cannot infer any additional information about $(W_k)_{k\in[K]}$ beyond that contained in the desired sum and known from the colluding users. That is, the following security constraint must be satisfied for any $\mathcal{T}$, where $\mathcal{T} \subset [K], |\mathcal{T}| \le T$.

[Security]

$$I\left((W_k)_{k\in[K]}; (X_k)_{k\in[K]} \middle| \sum_{k\in[K]} W_k, (W_k, Z_k)_{k\in\mathcal{T}}\right) = 0. \quad (6)$$

The communication *rate* R characterizes how many symbols each message contains per input symbol, and is defined as follows.

$$R \triangleq \frac{L_X}{L}. \quad (7)$$

The individual (total) key *rate* R_Z (R_{Z_Σ}) characterizes how many symbols each key variable (the source key variable) contains per input symbol, and is defined as follows.

$$R_Z \triangleq \frac{L_Z}{L}, \ R_{Z_\Sigma} \triangleq \frac{L_{Z_\Sigma}}{L}. \quad (8)$$

The rate tuple (R, R_Z, R_{Z_Σ}) is said to be achievable if there exists a secure summation scheme, for which the correctness constraint (5) and the security constraint (6) are satisfied, and the communication rate, the individual key rate, and the 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 (i.e., capacity region), denoted as $\mathcal{R}^*$.

A. General Groupwise Keys and Colluding Users

Consider the general groupwise key setting, i.e., any subset $\mathcal{G} \subset [K]$ of users may share an independent key variable $S_\mathcal{G}$ of size L_S symbols[3] from $\mathbb{F}_q$. Denote the family (set) of all sets of users that share a key by $\overline{\mathcal{G}}$. Then we can represent such groupwise keys by a hypergraph $H = (\mathcal{V}, \mathcal{E})$, defined as follows. The node set $\mathcal{V} = \{v_1, \cdots, v_K\}$ contains K nodes, where $v_k, k \in [K]$ represents User k; the edge set $\mathcal{E} = \{e_1, \cdots, e_{|\overline{\mathcal{G}}|}\}$ contains $|\overline{\mathcal{G}}|$ (hyper)edges, where each edge[4] $e_i, i \in [|\overline{\mathcal{G}}|]$ represents a group of users that share a same key, i.e., $e_i = \{(v_j)_{j\in\mathcal{G}_i}\}$ where $\mathcal{G}_i$ is the i-th element of $\overline{\mathcal{G}}$ (we may take any order of the elements in the set). A key hypergraph example can be found in Fig. 2(a), where $\overline{\mathcal{G}} = \{\{1,2,4\},\{2,3\},\{3,4\}\}$ and $e_1 = \{v_1, v_2, v_4\}, e_2 = \{v_2, v_3\}, e_3 = \{v_3, v_4\}$. A hypergraph $H = (\mathcal{V}, \mathcal{E})$ (with at least two nodes) is said to be *connected* if for any $\mathcal{V}_1 \subset \mathcal{V}, 0 < |\mathcal{V}_1| < |\mathcal{V}|$, there exists at least one edge $e \in \mathcal{E}$ such that $e \not\subset \mathcal{V}_1$ and $e \not\subset \mathcal{V}\backslash\mathcal{V}_1$, i.e., for any two-part partition of the node set, there exists at least one edge that connects the two parts.

Similarly, we assume that any subset $\mathcal{T} \subset [K]$ of users may collude with the server. Denote the family of all colluding user set by $\overline{\mathcal{T}}$ and the security constraint is the same as that in (6), but it holds for all $\mathcal{T} \in \overline{\mathcal{T}}$. We will study the feasibility of secure summation, i.e., if the optimal rate region $\mathcal{R}^*_{gc}$ (closure of achievable rate tuples (R, R_S), where $R_S \triangleq L_S/L$) is empty or not. The condition turns out to be related to the connectivity of the key hypergraph H after removing a colluding user set and the known keys, which is denoted by $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$, i.e., the induced subgraph of H with node set $\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}$.

B. Symmetric Groupwise Keys and Colluding Users

The symmetric groupwise key setting is a special case of the general groupwise key setting presented in the previous section and refers to a specific type of joint distribution of the keys, where every $1 \le G \le K$ users share an equal-size independent key. Consider $\binom{K}{G}$ independent random variables $S_\mathcal{G}, \mathcal{G} \in \binom{[K]}{G}$ and each $S_\mathcal{G}$ is comprised of L_S i.i.d. uniform symbols from $\mathbb{F}_q$.

$$H\left((S_\mathcal{G})_{\mathcal{G}\in\binom{[K]}{G}}\right) = \sum_{\mathcal{G}\in\binom{[K]}{G}} H(S_\mathcal{G}) = \binom{K}{G} L_S. \quad (9)$$

For example, when $K = 4, G = 3$, we have $S_{\{1,2,3\}}$, $S_{\{1,2,4\}}$, $S_{\{1,3,4\}}$, $S_{\{2,3,4\}}$. The key variable $S_\mathcal{G}$ is shared by users in $\mathcal{G}$ so that the key variable held by User k, Z_k is given by

$$Z_k = (S_\mathcal{G})_{k\in\mathcal{G}, \mathcal{G}\in\binom{[K]}{G}}, \forall k \in [K]. \quad (10)$$

For example, when $K = 4$, $G = 3$, $Z_3 = (S_{\{1,2,3\}}, S_{\{1,3,4\}}, S_{\{2,3,4\}})$. The groupwise key rate R_S characterizes

[3]We are interested in the feasibility of the secure summation problem for the general groupwise key setting so that we may assume that each key has the same size with no loss.

[4]When the user set $\mathcal{G}$ contains only a single user, i.e., $|\mathcal{G}| = 1$, we may either assume there is a self-loop or no edge, both of which will not influence our result as it only depends on connectivity.

how many symbols each groupwise key variable contains per input symbol, and is defined as follows.

$$R_S \triangleq \frac{L_S}{L}. \tag{11}$$

Note that due to the symmetry assumption on the groupwise keys, the individual and total key rates R_Z, R_{Z_Σ} can be readily obtained from R_S, i.e., $R_Z = \binom{K-1}{G-1}R_S, R_{Z_\Sigma} = \binom{K}{G}R_S$. Thus for the symmetric groupwise key setting, the rate tuple (R, R_Z, R_{Z_Σ}) can be more succinctly captured by (R, R_S), which will be adopted. The colluding user set is assumed to be symmetric, i.e., any T users may collude with the server (refer to (6)). The closure of the set of all achievable rate tuples (R, R_S) is called the optimal rate region,[5] denoted as $\mathcal{R}_g^*$.

III. Results

In this section, we summarize our main results along with key observations.

A. Secure Summation: Capacity Region

The optimal rate region of secure summation is characterized in Theorem 1, presented below.

Theorem 1: For the secure summation problem with $K \geq 2$ users and at most $0 \leq T \leq K-2$ colluding users, the optimal rate region is

$$\mathcal{R}^* = \{(R, R_Z, R_{Z_\Sigma}) : R \geq 1, R_Z \geq 1, R_{Z_\Sigma} \geq K-1\}. \tag{12}$$

An intuitive explanation of Theorem 1 may be seen as follows. To securely compute 1 symbol of the sum at the server, each user needs to send at least 1 symbol (which carries its own input), each user needs to hold a key of at least 1 symbol (to protect the 1 symbol transmit message), and all users must hold some key variables of at least $K-1$ symbols (note that the server must be able to only decode the 1 symbol sum from all K symbols received, so the remaining $K-1$ symbols must be fully protected by some key variables). The proof of Theorem 1 is presented in Section IV.

B. Secure Summation With General Groupwise Keys and Colluding Users: Feasibility Condition

The feasibility condition of secure summation with general groupwise keys and colluding users is characterized in Theorem 2, presented below.

Theorem 2: For the secure summation problem with groupwise key hypergraph $H = (\mathcal{V}, \mathcal{E})$ and colluding user set family $\overline{\mathcal{T}}$,

$$\mathcal{R}_{gc}^* \neq \emptyset \text{ if and only if } H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}] \text{ is connected for any } \mathcal{T} \in \overline{\mathcal{T}}. \tag{13}$$

An intuitive explanation of Theorem 2 may be seen as follows. On the one hand, when the induced subgraph of H is not connected after removing some colluding user set, the remaining keys are not sufficiently correlated to protect the desired sum as there exist two sets of users (a two-part partition of the subgraph) whose keys are independent (note that there is no edge connecting the two parts, i.e., no key is known to both parts). On the other hand, when the induced subgraph of H is connected no matter which colluding user set is removed, we may fully use all existing groupwise keys to produce a secure summation scheme that is secure to all colluding user sets. The detailed proof of Theorem 2 is presented in Section V.

C. Secure Summation With Symmetric Groupwise Keys and Colluding Users: Capacity Region

The optimal rate region of secure summation with symmetric groupwise keys and colluding users is characterized in Theorem 3, presented below.

Theorem 3: For the secure summation problem with $K \geq 2$ users, at most $0 \leq T \leq K-2$ colluding users, and symmetric groupwise keys of group size $1 \leq G \leq K$, the optimal rate region is

$$\mathcal{R}_g^* = \begin{cases} \emptyset & \text{if } G = 1 \text{ or } G > K-T, \\ \left\{ (R, R_S) : R \geq 1, \right. & \\ \left. R_S \geq \frac{K-T-1}{\binom{K-T}{G}} \right\} & \text{else } 1 < G \leq K-T. \end{cases} \tag{14}$$

An intuitive explanation of Theorem 3 may be seen as follows. The $G = 1$ setting is not feasible as here the keys are uncorrelated; otherwise when $G > K-T$, consider any set of T colluding users $\mathcal{T}$, who know all the keys as any groupwise key $S_\mathcal{G}, |\mathcal{G}| = G$ involves at least one user in $\mathcal{T}$. As all the keys are known to the server (through user collusion), nothing can be hidden from the server, which violates the security constraint so that secure summation is not feasible. Note that the $G > K-T$ case of Theorem 3 is covered by Theorem 2 as the edge set of $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ with symmetric groupwise keys is empty as all the keys are known to the colluding users (thus $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is trivially not connected). Also, the $G = 1$ case is trivially covered by Theorem 2. When $1 < G \leq K-T$, suppose for now we have deleted T colluding users and the keys known to them, leaving us with a secure summation problem with $K-T$ users and symmetric groupwise keys of size G among them. For this problem with $K-T$ users, from Theorem 1, the total key size should be at least $K-T-1$, i.e., $K-T-1 \leq \binom{K-T}{G}R_S$ and we have the desired converse on the key rate. Guided by the converse, we will design a vector linear scheme, where the keys are precoded by matrices with proper size such that the sum of the keys from a given group is equal to zero (to guarantee correctness) and the overall key blocks are sufficiently generic (i.e., full rank, to guarantee security). The details are deferred to the proof of Theorem 3, presented in Section VI.

[5] Some explanation on the notation - $\mathcal{R}^*$ denotes the optimal rate region with arbitrarily coded keys, $\mathcal{R}_{gc}^*$ denotes the optimal rate region with general groupwise keys and colluding user sets, and $\mathcal{R}_g^*$ denotes the optimal rate region with symmetric groupwise keys and colluding users.

IV. Proof of Theorem 1

A. Converse

We start with a few useful lemmas. First, we show that each X_k must contain at least L symbols (the size of the input) even if all other inputs are known.

Lemma 1: For any $u \in [K]$, we have

$$H\left(X_u|(W_k, Z_k)_{k\in[K]\backslash\{u\}}\right) \geq L. \tag{15}$$

Proof:

$$\begin{aligned}
&H\left(X_u|(W_k, Z_k)_{k\in[K]\backslash\{u\}}\right)\\
&\geq I\left(X_u; \sum_{k\in[K]} W_k \middle| (W_k, Z_k)_{k\in[K]\backslash\{u\}}\right) &(16)\\
&= H\left(\sum_{k\in[K]} W_k \middle| (W_k, Z_k)_{k\in[K]\backslash\{u\}}\right)\\
&\quad - H\left(\sum_{k\in[K]} W_k \middle| X_u, (W_k, Z_k)_{k\in[K]\backslash\{u\}}\right) &(17)\\
&\overset{(1)(4)}{\geq} H\left(W_u\right) - H\left(\sum_{k\in[K]} W_k \middle| (X_k)_{k\in[K]}\right) &(18)\\
&\overset{(2)(5)}{=} L &(19)
\end{aligned}$$

where the first term of (18) follows from the fact that input W_u is independent of other inputs and keys $(W_k, Z_k)_{k\in[K]\backslash\{u\}}$ (see (1)) and the second term of (18) follows from the fact that $(X_k)_{k\in[K]\backslash\{u\}}$ is determined by $(W_k, Z_k)_{k\in[K]\backslash\{u\}}$ (see (4)). In (19), we use the property that W_u has L uniform symbols (see (2)) and the desired sum $\sum_{k\in[K]} W_k$ can be decoded with no error from all messages $(X_k)_{k\in[K]}$ (see (5)). ∎

Second, we show that the messages from any set of users must contain all information about their input sum (i.e., L symbols).

Lemma 2: For any set of colluding users $\mathcal{T}$, denote its complement as $\mathcal{T}^c \triangleq [K]\backslash\mathcal{T}$ and we have

$$I\left((X_k)_{k\in\mathcal{T}^c}; (W_k)_{k\in\mathcal{T}^c}|(W_k, Z_k)_{k\in\mathcal{T}}\right) = L. \tag{20}$$

Proof:

$$\begin{aligned}
&I\left((X_k)_{k\in\mathcal{T}^c}; (W_k)_{k\in\mathcal{T}^c}|(W_k, Z_k)_{k\in\mathcal{T}}\right)\\
&= I\left((X_k)_{k\in\mathcal{T}^c}; (W_k)_{k\in\mathcal{T}^c}, \sum_{k\in\mathcal{T}^c} W_k \middle| (W_k, Z_k)_{k\in\mathcal{T}}\right) &(21)\\
&= I\left((X_k)_{k\in\mathcal{T}^c}; \sum_{k\in\mathcal{T}^c} W_k \middle| (W_k, Z_k)_{k\in\mathcal{T}}\right)\\
&\quad + \underbrace{I\left((X_k)_{k\in\mathcal{T}^c}; (W_k)_{k\in\mathcal{T}^c} \middle| \sum_{k\in\mathcal{T}^c} W_k, (W_k, Z_k)_{k\in\mathcal{T}}\right)}_{\overset{(6)}{=}0} &(22)\\
&= H\left(\sum_{k\in\mathcal{T}^c} W_k \middle| (W_k, Z_k)_{k\in\mathcal{T}}\right)\\
&\quad - \underbrace{H\left(\sum_{k\in\mathcal{T}^c} W_k \middle| (X_k)_{k\in\mathcal{T}^c}, (W_k, Z_k)_{k\in\mathcal{T}}\right)}_{\overset{(4)(5)}{=}0} &(23)\\
&\overset{(1)(2)}{=} L &(24)
\end{aligned}$$

where the last step follows from the independence and uniformity of the inputs. ∎

Third, we show that the keys only known to non-colluding users must be sufficient large to protect their inputs (beyond the desired sum).

Lemma 3: For any set of colluding users $\mathcal{T}$ and its complement $\mathcal{T}^c = [K]\backslash\mathcal{T}$, we have

$$H((Z_k)_{k\in\mathcal{T}^c}|(Z_k)_{k\in\mathcal{T}}) \geq (K - |\mathcal{T}| - 1)L. \tag{25}$$

Proof: First, we use Lemma 1 to obtain

$$\begin{aligned}
&H\left((X_k)_{k\in\mathcal{T}^c}|(W_k)_{k\in\mathcal{T}}, (Z_k)_{k\in\mathcal{T}}\right)\\
&\geq \sum_{u\in\mathcal{T}^c} H\left(X_u|(W_k, Z_k)_{k\in[K]\backslash\{u\}}\right) &(26)\\
&\overset{(15)}{\geq} (K - |\mathcal{T}|)L &(27)
\end{aligned}$$

where in (26), we use the chain rule and the property that conditioning cannot increase entropy. Next,

$$\begin{aligned}
&H((Z_k)_{k\in\mathcal{T}^c}|(Z_k)_{k\in\mathcal{T}})\\
&\geq I\left((Z_k)_{k\in\mathcal{T}^c}; (X_k)_{k\in\mathcal{T}^c}|(W_k)_{k\in\mathcal{T}^c}, (Z_k)_{k\in\mathcal{T}}\right) &(28)\\
&\overset{(4)}{=} H\left((X_k)_{k\in\mathcal{T}^c}|(W_k)_{k\in\mathcal{T}^c}, (Z_k)_{k\in\mathcal{T}}\right) &(29)\\
&\geq H\left((X_k)_{k\in\mathcal{T}^c}|(W_k)_{k\in[K]}, (Z_k)_{k\in\mathcal{T}}\right) &(30)\\
&= H\left((X_k)_{k\in\mathcal{T}^c}|(W_k)_{k\in\mathcal{T}}, (Z_k)_{k\in\mathcal{T}}\right)\\
&\quad - I\left((X_k)_{k\in\mathcal{T}^c}; (W_k)_{k\in\mathcal{T}^c}|(W_k)_{k\in\mathcal{T}}, (Z_k)_{k\in\mathcal{T}}\right) &(31)\\
&\overset{(27)(20)}{\geq} (K - |\mathcal{T}|)L - L = (K - |\mathcal{T}| - 1)L. &(32)
\end{aligned}$$

∎

We are now ready to prove the converse of Theorem 1.

Proof of $R \geq 1$: Consider any user $u \in [K]$.

$$\begin{aligned}
L_X &\geq H\left(X_u\right) &(33)\\
&\geq H\left(X_u|(W_k, Z_k)_{k\in[K]\backslash\{u\}}\right) &(34)\\
&\overset{(15)}{\geq} L &(35)\\
\Rightarrow \quad R &\overset{(7)}{=} \frac{L_X}{L} \geq 1. &(36)
\end{aligned}$$

Proof of $R_Z \geq 1$: First, we show that the message $X_u, u \in [K]$ is independent of the input W_u.

$$I\left(X_u; W_u\right)$$

$$\leq I\left(X_u, \sum_{k\in[K]} W_k; W_u\right) \tag{37}$$

$$= I\left(\sum_{k\in[K]} W_k; W_u\right) + I\left(X_u; W_u \middle| \sum_{k\in[K]} W_k\right) \tag{38}$$

$$\overset{(1)(2)}{\leq} I\left((W_k)_{k\in[K]}; (X_k)_{k\in[K]} \middle| \sum_{k\in[K]} W_k\right) \tag{39}$$

$$\overset{(6)}{=} 0 \tag{40}$$

where (39) follows from the observation that the first term of (38) is zero because $K \geq 2$ and the inputs are independent and uniform. To obtain (40), we set $\mathcal{T} = \emptyset$ in the security constraint (6). Next, consider any $u \in [K]$.

$$L_Z \geq H(Z_u) \tag{41}$$

$$\geq I(Z_u; X_u|W_u) \tag{42}$$

$$\overset{(4)}{=} H(X_u|W_u) \tag{43}$$

$$\overset{(40)}{=} H(X_u) \tag{44}$$

$$\overset{(35)}{\geq} L \tag{45}$$

$$\Rightarrow \quad R_Z \overset{(8)}{=} \frac{L_Z}{L} \geq 1. \tag{46}$$

Proof of $R_{Z_\Sigma} \geq K-1$*:* Set $\mathcal{T} = \emptyset$ so that its complement $\mathcal{T}^c = [K]$.

$$L_{Z_\Sigma} \geq H(Z_\Sigma) \tag{47}$$

$$\overset{(3)}{\geq} H\left((Z_k)_{k\in[K]}\right) \tag{48}$$

$$\geq H((Z_k)_{k\in\mathcal{T}^c}|(Z_k)_{k\in\mathcal{T}}) \tag{49}$$

$$\overset{(25)}{\geq} (K-|\mathcal{T}|-1)L = (K-1)L \tag{50}$$

$$\Rightarrow \quad R_{Z_\Sigma} \overset{(8)}{=} \frac{L_{Z_\Sigma}}{L} \geq K-1 \tag{51}$$

where in (48), we use the fact that all key variables $(Z_k)_{k\in[K]}$ are generated from the source key variable Z_Σ (see (3)). In (50), we use (25) in Lemma 3.

B. Achievability

The achievable scheme is straightforward and is plotted in Fig. 1. The only non-trivial aspect is the proof of security, presented below.

We first describe the scheme. Consider $K-1$ i.i.d. uniform variables over $\mathbb{F}_q$, $N_1, \cdots, N_{K-1}$ and set the key variables as

$$Z_k = N_k, \forall k \in [K-1]$$

$$Z_K = -N_1 - \cdots - N_{K-1} \triangleq N_K. \tag{52}$$

Set $L = 1$ and set the messages as

$$X_k = W_k + N_k, \forall k \in [K]. \tag{53}$$

Note that $L_X = 1, L_Z = 1, L_{Z_\Sigma} = K-1$, so the rate achieved is $R = L_X/L = 1, R_Z = L_Z/L = 1, R_{Z_\Sigma} = L_{Z_\Sigma}/L = K-1$, as desired. Correctness is proved by noting that $\sum_{k\in[K]} X_k = \sum_{k\in[K]} W_k$. We are left to prove the security. For any colluding user set $\mathcal{T}$, we verify that the security constraint (6) is satisfied.

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

$$\overset{(52)(53)}{=} I\left((W_k)_{k\in\mathcal{T}^c}; (W_k+N_k)_{k\in\mathcal{T}^c} \middle| \cdots \right.$$

$$\left. \cdots \sum_{k\in\mathcal{T}^c} W_k, (W_k, N_k)_{k\in\mathcal{T}}\right) \tag{54}$$

$$= H\left((W_k+N_k)_{k\in\mathcal{T}^c} \middle| \sum_{k\in\mathcal{T}^c} W_k, (W_k, N_k)_{k\in\mathcal{T}}\right)$$

$$- H\left((W_k+N_k)_{k\in\mathcal{T}^c} \middle| \sum_{k\in\mathcal{T}^c} W_k, \dots \right.$$

$$\left. \dots (W_k, N_k)_{k\in\mathcal{T}}, (W_k)_{k\in\mathcal{T}^c}\right) \tag{55}$$

$$\leq H\left((W_k+N_k)_{k\in\mathcal{T}^c} \middle| \sum_{k\in\mathcal{T}^c} (W_k+N_k)\right)$$

$$- H\left((N_k)_{k\in\mathcal{T}^c} | (N_k)_{k\in\mathcal{T}}\right) \tag{56}$$

$$\leq (K-|\mathcal{T}|-1) - (K-|\mathcal{T}|-1) = 0 \tag{57}$$

where in (54), we plug in the message and key variable assignment (refer to (52), (53)). In (56), the first term follows from fact that $0 = \sum_{k\in[K]} N_k = \sum_{k\in\mathcal{T}} N_k + \sum_{k\in\mathcal{T}^c} N_k$ (refer to (52)), i.e., $\sum_{k\in\mathcal{T}^c} N_k$ can be obtained from $(N_k)_{k\in\mathcal{T}}$ and the property that dropping conditioning cannot reduce entropy; the second term is obtained by applying the independence of the input and key variables (refer to (1)). In the last step, the first term follows from the fact that the first term contains at most $K-|\mathcal{T}|-1$ terms after conditioning and uniform distribution maximizes entropy; the second term follows from the independence and uniformity of the $N_1, \cdots, N_{K-1}$ variables and the fact that $\sum_{k\in\mathcal{T}^c} N_k$ can be obtained from $(N_k)_{k\in\mathcal{T}}$.

V. Proof of Theorem 2

The proof of Theorem 2 is comprised of two directions. On the one hand, we show that $\mathcal{R}_{gc}^* \neq \emptyset \Rightarrow H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is connected for any $\mathcal{T} \in \overline{\mathcal{T}}$, i.e., $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is not connected for some $\mathcal{T} \in \overline{\mathcal{T}} \Rightarrow \mathcal{R}_{gc}^* = \emptyset$, which requires a converse proof. On the other hand, we show that $\mathcal{R}_{gc}^* \neq \emptyset \Leftarrow H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is connected for any $\mathcal{T} \in \overline{\mathcal{T}}$, which requires an achievability proof. These two proofs are provided next.

A. Converse: $\exists \mathcal{T}, H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ Is Not Connected $\Rightarrow \mathcal{R}^*_{gc}=\emptyset$

As $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is not connected, by definition, there exists a two-part partition of the nodes, $\mathcal{V}_1, \mathcal{V}_2$, where $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}, |\mathcal{V}_1| \geq 1, |\mathcal{V}_2| \geq 1$ such that any edge e of $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ must fully lie in $\mathcal{V}_1$ or $\mathcal{V}_2$, i.e., $e \subset \mathcal{V}_1$ or $e \subset \mathcal{V}_2$. Denote the set of the indices of the nodes in $\mathcal{V}_1$ and $\mathcal{V}_2$ (i.e., the set of users) by $\mathcal{U}_1$ and $\mathcal{U}_2$, respectively. Then $\mathcal{U}_1 \cup \mathcal{U}_2 = [K]\backslash\mathcal{T}, |\mathcal{U}_1| \geq 1, |\mathcal{U}_2| \geq 1$ and

$$I\left((Z_k)_{k\in\mathcal{U}_1}; (Z_k)_{k\in\mathcal{U}_2}\middle|(Z_k)_{k\in\mathcal{T}}\right) \overset{(9)}{=} 0 \tag{58}$$

because Z_k corresponds to all edges incident with node v_k in the hypergraph H; conditioning on $(Z_k)_{k\in\mathcal{T}}$ corresponds to the removal of the nodes $\{(v_k)_{k\in\mathcal{T}}\}$ and the incident edges (keys); all remaining edges e fully lying in $\mathcal{V}_1$ or $\mathcal{V}_2$ leads to that all remaining keys are known either only to users from $\mathcal{U}_1$ or only to users from $\mathcal{U}_2$. Combining with the fact that the gorupwise keys are independent, we conclude that the above conditional mutual information term is zero.

Next, we show that the above mutual information term must be strictly positive, i.e., the keys must be correlated.

$$\begin{aligned}
&I\left((Z_k)_{k\in\mathcal{U}_1}; (Z_k)_{k\in\mathcal{U}_2}|(Z_k)_{k\in\mathcal{T}}\right)\\
&\overset{(1)}{=} I\left((W_k, Z_k)_{k\in\mathcal{U}_1}; (W_k, Z_k)_{k\in\mathcal{U}_2}|(W_k, Z_k)_{k\in\mathcal{T}}\right) &(59)\\
&\overset{(4)}{\geq} I\left((W_k, X_k)_{k\in\mathcal{U}_1}; (W_k, X_k)_{k\in\mathcal{U}_2}|(W_k, Z_k)_{k\in\mathcal{T}}\right) &(60)\\
&\geq I\left(\sum_{k\in\mathcal{U}_1} W_k; (W_k, X_k)_{k\in\mathcal{U}_2}\middle|(X_k)_{k\in\mathcal{U}_1}, (W_k, Z_k)_{k\in\mathcal{T}}\right) &(61)\\
&= H\left(\sum_{k\in\mathcal{U}_1} W_k\middle|(X_k)_{k\in\mathcal{U}_1}, (W_k, Z_k)_{k\in\mathcal{T}}\right)\\
&\quad \underbrace{- H\left(\sum_{k\in\mathcal{U}_1} W_k\middle|(X_k)_{k\in\mathcal{U}_1}, (W_k, X_k)_{k\in\mathcal{U}_2}, (W_k, Z_k)_{k\in\mathcal{T}}\right)}_{\overset{(4)(5)}{=}0} &(62)\\
&= H\left(\sum_{k\in\mathcal{U}_1} W_k\middle|(W_k, Z_k)_{k\in\mathcal{T}}\right)\\
&\quad \underbrace{- I\left(\sum_{k\in\mathcal{U}_1} W_k; (X_k)_{k\in\mathcal{U}_1}\middle|(W_k, Z_k)_{k\in\mathcal{T}}\right)}_{\overset{(6)}{\leq}0} &(63)\\
&\overset{(1)(2)}{\geq} L &(64)
\end{aligned}$$

where in (59), we use the independence of the input and key variables and the fact that $\mathcal{U}_1, \mathcal{U}_2, \mathcal{T}$ are disjoint. The second term of (62) is zero because $\mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{T} = [K]$ and $\sum_{k\in[K]} W_k$ can be recovered from $(X_k)_{k\in[K]}$. The second term of (63) is zero due to the security constraint (6). The last step follows from the independence and uniformity of the inputs, and $\mathcal{U}_1 \cap \mathcal{T} = \emptyset$.

Comparing (64) with (58), we arrive at the contradiction (i.e., $L \leq 0$) and complete the proof that $\mathcal{R}^*_{gc} = \emptyset$ (i.e., secure summation is not feasible).

B. Achievability: $\forall \mathcal{T}, H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is connected $\Rightarrow \mathcal{R}^*_{gc} \neq \emptyset$

We first give an achievable scheme that uses all available groupwise keys (in the same zero-sum manner as in Theorem 3) and then prove that the scheme is correct and secure.

Suppose the key hypergraph $H = (\mathcal{V}, \mathcal{E})$ contains K nodes $v_1, \cdots, v_K$ and $\left|\overline{\mathcal{G}}\right|$ edges $e_1, \cdots, e_{\left|\overline{\mathcal{G}}\right|}$, where $\overline{\mathcal{G}}$ is the family of the sets of users that share an independent key, i.e., $\overline{\mathcal{G}} = \{\mathcal{G}_1, \cdots, \mathcal{G}_{\left|\overline{\mathcal{G}}\right|}\}$. Then we have $\left|\overline{\mathcal{G}}\right|$ groupwise keys and suppose the key shared by users in $\mathcal{G}_i, \forall i \in [\left|\overline{\mathcal{G}}\right|]$, $S_{\mathcal{G}_i} \in \mathbb{F}_q^{(|\mathcal{G}_i|-1)\times 1}$ has length[6] $|\mathcal{G}_i| - 1$.

Set $L = L_X = 1$ and the messages as

$$X_k = W_k + \sum_{\mathcal{G}: k\in\mathcal{G}, \mathcal{G}\in\overline{\mathcal{G}}} \mathbf{h}^k_{\mathcal{G}} S_{\mathcal{G}}, \ \forall k \in [K] \tag{65}$$

where $\mathbf{h}^k_{\mathcal{G}}$ is a $1 \times (|\mathcal{G}| - 1)$ vector set as (suppose $\mathcal{G} = \{u_1, \cdots, u_{|\mathcal{G}|}\} \subset [K]$)

$$\begin{aligned}
\mathbf{h}^{u_1}_{\mathcal{G}} &= [1, 0, \cdots, 0]\\
\mathbf{h}^{u_2}_{\mathcal{G}} &= [0, 1, \cdots, 0]\\
&\vdots\\
\mathbf{h}^{u_{|\mathcal{G}|-1}}_{\mathcal{G}} &= [0, 0, \cdots, 1]\\
\mathbf{h}^{u_{|\mathcal{G}|}}_{\mathcal{G}} &= [-1, -1, \cdots, -1],
\end{aligned} \tag{66}$$

so

$$\sum_{k\in\mathcal{G}} \mathbf{h}^k_{\mathcal{G}} = 0. \tag{67}$$

Note that for each group $\mathcal{G} \in \overline{\mathcal{G}}$, the sum of all precoded key variables in the messages is zero, so correctness is guaranteed.

$$\begin{aligned}
\sum_{k\in[K]} X_k &= \sum_{k\in[K]} W_k + \sum_{k\in[K]} \sum_{\mathcal{G}: k\in\mathcal{G}, \mathcal{G}\in\overline{\mathcal{G}}} \mathbf{h}^k_{\mathcal{G}} S_{\mathcal{G}} &(68)\\
&= \sum_{k\in[K]} W_k + \sum_{\mathcal{G}: \mathcal{G}\in\overline{\mathcal{G}}} \left[\left(\sum_{k\in\mathcal{G}} \mathbf{h}^k_{\mathcal{G}}\right) S_{\mathcal{G}}\right] &(69)\\
&\overset{(67)}{=} \sum_{k\in[K]} W_k. &(70)
\end{aligned}$$

Finally, we prove that the scheme (65) is secure when $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is connected for any $\mathcal{T}$. Consider the security constraint (6) for any colluding set $\mathcal{T} \in \overline{\mathcal{T}}$.

$$I\left((W_k)_{k\in[K]\backslash\mathcal{T}}; (X_k)_{k\in[K]\backslash\mathcal{T}}\middle|\cdots\right.$$

[6]We will use $|\mathcal{G}_i| - 1$ symbols from $S_{\mathcal{G}_i}$. If we want to make each key has the same length L_S, we can set L_S as $\max_i(|\mathcal{G}_i| - 1)$ and zero pad shorter keys.

$$\cdots \sum_{k\in[K]\backslash\mathcal{T}} W_k, (W_k, Z_k)_{k\in\mathcal{T}}\Bigg) \tag{71}$$

$$= H\left((X_k)_{k\in[K]\backslash\mathcal{T}} \Bigg| \sum_{k\in[K]\backslash\mathcal{T}} W_k, (W_k, Z_k)_{k\in\mathcal{T}}\right) - H\left((X_k)_{k\in[K]\backslash\mathcal{T}} \Big| (W_k)_{k\in[K]}, (Z_k)_{k\in\mathcal{T}}\right) \tag{72}$$

$$\le (K-|\mathcal{T}|-1) - H\left(\left(\sum_{\mathcal{G}:k\in\mathcal{G},\mathcal{G}\in\overline{\mathcal{G}}} \mathbf{h}_{\mathcal{G}}^k S_{\mathcal{G}}\right)_{k\in[K]\backslash\mathcal{T}} \Big| (W_k)_{k\in[K]}, (Z_k)_{k\in\mathcal{T}}\right) \tag{73}$$

$$= (K-|\mathcal{T}|-1) - H\left(\left(\sum_{\mathcal{G}:k\in\mathcal{G},\mathcal{G}\in\overline{\mathcal{G}},\mathcal{G}\cap\mathcal{T}=\emptyset} \mathbf{h}_{\mathcal{G}}^k S_{\mathcal{G}}\right)_{k\in[K]\backslash\mathcal{T}}\right) \tag{74}$$

$$= (K-|\mathcal{T}|-1) - (K-|\mathcal{T}|-1) = 0 \tag{75}$$

where the last step relies on the generic property of the preocoded keys and the connectivity property of $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$, and is derived as follows. Suppose there are M groupwise keys that are only known to non-colluding users in $[K]\backslash\mathcal{T}$ and denote the corresponding set of users as $\mathcal{G}_{j_1}, \cdots, \mathcal{G}_{j_M}$, i.e., $\forall m \in [M], \mathcal{G}_{j_m} \in \overline{\mathcal{G}}, \mathcal{G}_{j_m} \cap \mathcal{T} = \emptyset$. Denote $[K]\backslash\mathcal{T} = \{u_1, \cdots, u_{K-|\mathcal{T}|}\}$. Then

$$H\left(\left(\sum_{\mathcal{G}:k\in\mathcal{G},\mathcal{G}\in\overline{\mathcal{G}},\mathcal{G}\cap\mathcal{T}=\emptyset} \mathbf{h}_{\mathcal{G}}^k S_{\mathcal{G}}\right)_{k\in[K]\backslash\mathcal{T}}\right)$$

$$= H\left(\underbrace{\begin{bmatrix} \mathbf{h}_{\mathcal{G}_{j_1}}^{u_1} & \mathbf{h}_{\mathcal{G}_{j_2}}^{u_1} & \cdots & \mathbf{h}_{\mathcal{G}_{j_M}}^{u_1} \\ \mathbf{h}_{\mathcal{G}_{j_1}}^{u_2} & \mathbf{h}_{\mathcal{G}_{j_2}}^{u_2} & \cdots & \mathbf{h}_{\mathcal{G}_{j_M}}^{u_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{\mathcal{G}_{j_1}}^{u_{K-|\mathcal{T}|}} & \mathbf{h}_{\mathcal{G}_{j_2}}^{u_{K-|\mathcal{T}|}} & \cdots & \mathbf{h}_{\mathcal{G}_{j_M}}^{u_{K-|\mathcal{T}|}} \end{bmatrix}}_{=\hat{\mathbf{H}}} \begin{bmatrix} S_{\mathcal{G}_{j_1}} \\ \vdots \\ S_{\mathcal{G}_{j_M}} \end{bmatrix}\right) \tag{76}$$

$$= \text{rank}(\hat{\mathbf{H}}) = K-|\mathcal{T}|-1 \tag{77}$$

where $\mathbf{h}_{\mathcal{G}}^u = \mathbf{0}$ if $u \notin \mathcal{G}$ and otherwise $\mathbf{h}_{\mathcal{G}}^u$ is specified in (66). The rank of $\hat{\mathbf{H}}$ is $K-|\mathcal{T}|-1$ because the first $K-|\mathcal{T}|-1$ rows of $\hat{\mathbf{H}}$ are linearly independent (and the last row is the sum of all above rows). Suppose otherwise, i.e., some subsets of the first $K-|\mathcal{T}|-1$ rows are linearly dependent. Denote the set of the indices of such rows by $\mathcal{G}'$. Then due to the assignment (66), we know that for any column block with subscript $\mathcal{G}_{j_m}$, we must have all $|\mathcal{G}_{j_m}|$ vectors $\mathbf{h}_{\mathcal{G}_{j_m}}^u, u \in \mathcal{G}_{j_m}$ to produce linearly dependent rows and it must hold that $\forall m \in [M], \mathcal{G}_{j_m} \subset \mathcal{G}'$ or $\mathcal{G}_{j_m} \subset [K]\backslash(\mathcal{T}\cup\mathcal{G}')$. Noting that each $\mathcal{G}_{j_m}$ corresponds to an edge e_{j_m}, the above claim means that all edges fully belong to one part of a two-part partition of $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ (whose node sets are given by $(v_k)_{k\in\mathcal{G}'}$ and $(v_k)_{k\in[K]\backslash(\mathcal{T}\cup\mathcal{G}')}$), which violates the condition that the induced subgraph $H[\mathcal{V}\backslash\{(v_k)_{k\in\mathcal{T}}\}]$ is connected. We have thus arrived at the contradiction and have proved that $\text{rank}(\hat{\mathbf{H}}) = K-|\mathcal{T}|-1$.

As the scheme (65) is both correct and secure and the rate achieved is non-zero, we have proved that $\mathcal{R}_{gc}^* \neq \emptyset$.

Remark 1: The achievable schemes presented in this paper are built on the same idea of using zero-sum randomness (widely used in secure computation literature) and thus are intimately related. For example, the achievable scheme of Theorem 2 uses the basic scheme of Theorem 1 for all groups and the achievable scheme of Theorem 3 generalizes the scalar basic scheme of Theorem 1 to a vector form with further desgin in the precoding matrices, to be presented in the next section.

VI. Proof of Theorem 3

A. Converse

When $G = 1$, the key hypergraph is obviously not connected and Theorem 2 shows $\mathcal{R}_g^* = \emptyset$; otherwise when $G > K-T$, consider any set of colluding users $\mathcal{T}$, where $|\mathcal{T}| = T \le K-2$. Each groupwise key $S_{\mathcal{G}}, \forall \mathcal{G} \subset [K], |\mathcal{G}| = G$ is known to the colluding users so that the key hypergraph has no edges and has at least two nodes after removing the colluding users and we may invoke Theorem 2 to establish that $\mathcal{R}_g^* = \emptyset$. Henceforth, it suffices to consider the case when $1 < G \le K-T$.

Proof of $R \ge 1$: It follows from the proof of Theorem 1 as groupwise keys are a special case of arbitrarily coded keys.

Proof of $R_S \ge \frac{K-T-1}{\binom{K-T}{G}}$: Note that Lemma 3 holds for groupwise keys. Applying (25) to a colluding user set $\mathcal{T}$ such that $|\mathcal{T}| = T$, we have

$$(K-T-1)L$$

$$\overset{(25)}{\le} H\left((Z_k)_{k\in\mathcal{T}^c} | (Z_k)_{k\in\mathcal{T}}\right) \tag{78}$$

$$\overset{(10)}{=} H\left((S_{\mathcal{G}})_{\mathcal{G}\in\binom{[K]}{G},\mathcal{G}\cap\mathcal{T}^c\neq\emptyset} \Big| (S_{\mathcal{G}})_{\mathcal{G}\in\binom{[K]}{G},\mathcal{G}\cap\mathcal{T}\neq\emptyset}\right) \tag{79}$$

$$\overset{(9)}{=} H\left((S_{\mathcal{G}})_{\mathcal{G}\in\binom{[K]}{G},\mathcal{G}\cap\mathcal{T}^c\neq\emptyset,\mathcal{G}\cap\mathcal{T}=\emptyset}\right) \tag{80}$$

$$= H\left((S_{\mathcal{G}})_{\mathcal{G}\in\binom{\mathcal{T}^c}{G}}\right) \tag{81}$$

$$\overset{(9)}{=} \binom{K-T}{G} \times L_S \tag{82}$$

$$\Rightarrow \quad R_S \overset{(11)}{=} \frac{L_S}{L} \ge \frac{K-T-1}{\binom{K-T}{G}} \tag{83}$$

where in (79), we replace each key variable Z_k by the groupwise keys (refer to (10)) and in (80) and (83), we apply the independence of the groupwise keys (refer to (9)).

Remark 2: Note that we may apply the individual rate R_Z bound from Theorem 1 to the groupwise key setting to obtain a bound on the groupwise key rate R_S, which turns out to be not tight.

B. Achievability

We first present the achievable scheme for two examples, to illustrate the idea in a simpler setting.

1) Example: $K = 3, T = 0, G = 2$: Consider $K = 3$ users, where no user may collude with the server ($T = 0$). As the group size $G = 2$, we have 3 groupwise keys $S_{\{1,2\}}, S_{\{1,3\}}, S_{\{2,3\}}$.

The achievable scheme is based on interpreting the scheme from Section IV-B of Theorem 1 in the groupwise key setting and permuting it for symmetrization. Specifically, consider the following basic component of the secure summation scheme from Theorem 1.

$$X_1 = W_1 + A \tag{84}$$

$$X_2 = W_2 + B \tag{85}$$

$$X_3 = W_3 - A - B \tag{86}$$

where each X_k, W_k contains 1 symbol from $\mathbb{F}_q$ and A, B are two i.i.d. uniform key symbols from $\mathbb{F}_q$. Note that the above keys can be assigned through setting groupwise keys $S_{\{1,3\}} = A, S_{\{2,3\}} = B$. Correctness and security follow immediately from Theorem 1. To produce a symmetric scheme where each groupwise key has the same size (note that $S_{\{1,2\}}$ has not been used), we apply the above scheme to all permutation of the users $\{1, 2, 3\}$. In particular, we set $L = 3! = 6$ and each input has 6 symbols, i.e., $W_k = [W_k(1), \cdots, W_k(6)]^T$. Correspondingly, uniform key variables A, B have length 6 each. The message and key assignment is as follows.

$$\begin{aligned}
&X_1(1) = W_1(1) + A(1)\\
&X_2(1) = W_2(1) + B(1)\\
&X_3(1) = W_3(1) - A(1) - B(1)\\
&X_1(2) = W_1(2) + B(2)\\
&X_2(2) = W_2(2) - A(2) - B(2)\\
&X_3(2) = W_3(2) + A(2)\\
&X_1(3) = W_1(3) - A(3) - B(3)\\
&X_2(3) = W_2(3) + A(3)\\
&X_3(3) = W_3(3) + B(3)\\
&X_1(4) = W_1(4) + A(4)\\
&X_2(4) = W_2(4) - A(4) - B(4)\\
&X_3(4) = W_3(4) + B(4)\\
&X_1(5) = W_1(5) + B(5)\\
&X_2(5) = W_2(5) + A(5)\\
&X_3(5) = W_3(5) - A(5) - B(5)\\
&X_1(6) = W_1(6) - A(6) - B(6)\\
&X_2(6) = W_2(6) + B(6)\\
&X_3(6) = W_3(6) + A(6)
\end{aligned}$$

and

$$S_{\{1,2\}} = \begin{bmatrix} B(2)\\ A(3)\\ A(4)\\ B(6)\end{bmatrix} \quad S_{\{1,3\}} = \begin{bmatrix} A(1)\\ B(3)\\ B(5)\\ A(6)\end{bmatrix} \quad S_{\{2,3\}} = \begin{bmatrix} B(1)\\ A(2)\\ B(4)\\ A(5)\end{bmatrix} \tag{87}$$

where $L_X = 6, L_S = 4$ so that the rate achieved is $R = L_X/L = 1, R_S = L_S/L = 2/3 = (3-1)/\binom{3}{2}$, as desired.

Permutation retains the correctness and security of the scheme (note that the input and key variables for each permutation are independent so that the mutual information terms (54) in the security proof tensorize).

To facilitate the presentation of the achievable scheme when $T > 0$, it is convenient to describe the scheme in matrix form. In particular, the above scheme can be equivalently written as

$$\begin{bmatrix} X_1\\ X_2\\ X_3\end{bmatrix} = \begin{bmatrix} W_1\\ W_2\\ W_3\end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{H}_{\{1,2\}} & \mathbf{H}_{\{1,3\}} & \mathbf{0}_{6\times 4}\\ -\mathbf{H}_{\{1,2\}} & \mathbf{0}_{6\times 4} & \mathbf{H}_{\{2,3\}}\\ \mathbf{0}_{6\times 4} & -\mathbf{H}_{\{1,3\}} & -\mathbf{H}_{\{2,3\}}\end{bmatrix}}_{\triangleq \mathbf{H}_{18\times 12}} \begin{bmatrix} S_{\{1,2\}}\\ S_{\{1,3\}}\\ S_{\{2,3\}}\end{bmatrix} \tag{88}$$

where

$$\mathbf{H}_{\{1,2\}} = \begin{bmatrix} 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}$$

$$\mathbf{H}_{\{1,3\}} = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}$$

$$\mathbf{H}_{\{2,3\}} = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{bmatrix}. \tag{89}$$

Now the scheme design reduces to the assignment of the key precoding matrix $\mathbf{H}$ and it turns out that security can be guaranteed by certain rank property of $\mathbf{H}$ (refer to Lemma 4).

2) Example: $K = 5, T = 2, G = 2$: Consider $K = 5$ users, where at most $T = 2$ users may collude with the server, and every $G = 2$ users share a groupwise key. The rate to be achieved is $R = 1, R_S = (K - T - 1)/\binom{K-T}{G} = 2/3$. To this end, set $L = L_X = 3$, i.e., $W_k = [W_k(1), W_k(2), W_k(3))]^T \in$

$\mathbb{F}_q^{3\times 1}, k \in [3]$; set $L_S = 2$, i.e., $S_{\mathcal{G}} = [S_{\mathcal{G}}(1), S_{\mathcal{G}}(2)]^T \in \mathbb{F}_q^{2\times 1}, \mathcal{G} \subset [3], |\mathcal{G}| = 2$. The messages are set as

$$\begin{aligned}
X_1 &= W_1 + \mathbf{H}_{\{1,2\}}^{3\times 2} S_{\{1,2\}} + \mathbf{H}_{\{1,3\}} S_{\{1,3\}} + \mathbf{H}_{\{1,4\}} S_{\{1,4\}} \\
&\quad + \mathbf{H}_{\{1,5\}} S_{\{1,5\}} \\
X_2 &= W_2 - \mathbf{H}_{\{1,2\}} S_{\{1,2\}} + \mathbf{H}_{\{2,3\}} S_{\{2,3\}} + \mathbf{H}_{\{2,4\}} S_{\{2,4\}} \\
&\quad + \mathbf{H}_{\{2,5\}} S_{\{2,5\}} \\
X_3 &= W_3 - \mathbf{H}_{\{1,3\}} S_{\{1,3\}} - \mathbf{H}_{\{2,3\}} S_{\{2,3\}} + \mathbf{H}_{\{3,4\}} S_{\{3,4\}} \\
&\quad + \mathbf{H}_{\{3,5\}} S_{\{3,5\}} \\
X_4 &= W_4 - \mathbf{H}_{\{1,4\}} S_{\{1,4\}} - \mathbf{H}_{\{2,4\}} S_{\{2,4\}} - \mathbf{H}_{\{3,4\}} S_{\{3,4\}} \\
&\quad + \mathbf{H}_{\{4,5\}} S_{\{4,5\}} \\
X_5 &= W_5 - \mathbf{H}_{\{1,5\}} S_{\{1,5\}} - \mathbf{H}_{\{2,5\}} S_{\{2,5\}} - \mathbf{H}_{\{3,5\}} S_{\{3,5\}} \\
&\quad - \mathbf{H}_{\{4,5\}} S_{\{4,5\}}
\end{aligned} \tag{90}$$

where each $\mathbf{H}_{\mathcal{G}} \in \mathbb{F}_q^{3\times 2}$ is a 3×2 key precoding matrix. Note that $\mathbf{H}_{\mathcal{G}}$ and $-\mathbf{H}_{\mathcal{G}}$ are used by every $|\mathcal{G}| = 2$ users (i.e., zero-sum-randomness) so that correctness is guaranteed, i.e., $\sum_{k\in[5]} X_k = \sum_{k\in[5]} W_k$. Security is ensured if the matrices $\mathbf{H}_{\mathcal{G}}$ are sufficiently generic. In particular, we will show in the general proof that if $\mathbf{H}_{\mathcal{G}}$ are randomly drawn from a sufficiently large field (note that the size q of the field $\mathbb{F}_q$ remains the same, but we can code over longer blocks by enlarging input size L so that we are operating over a larger extension field), then there must exist a matrix construction such that security holds. For this setting, suppose $q = 5$ and we may set

$$\begin{aligned}
\mathbf{H}_{\{1,2\}} &= \begin{bmatrix} 3 & 3 \\ 1 & 4 \\ 2 & 4 \end{bmatrix} & \mathbf{H}_{\{1,3\}} &= \begin{bmatrix} 2 & 1 \\ 0 & 4 \\ 0 & 1 \end{bmatrix} \\
\mathbf{H}_{\{1,4\}} &= \begin{bmatrix} 4 & 1 \\ 1 & 0 \\ 4 & 1 \end{bmatrix} & \mathbf{H}_{\{1,5\}} &= \begin{bmatrix} 3 & 4 \\ 2 & 2 \\ 1 & 2 \end{bmatrix} \\
\mathbf{H}_{\{2,3\}} &= \begin{bmatrix} 4 & 3 \\ 1 & 1 \\ 3 & 2 \end{bmatrix} & \mathbf{H}_{\{2,4\}} &= \begin{bmatrix} 0 & 3 \\ 0 & 4 \\ 2 & 0 \end{bmatrix} \\
\mathbf{H}_{\{2,5\}} &= \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 0 & 3 \end{bmatrix} & \mathbf{H}_{\{3,4\}} &= \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 0 & 3 \end{bmatrix} \\
\mathbf{H}_{\{3,5\}} &= \begin{bmatrix} 3 & 0 \\ 3 & 1 \\ 2 & 4 \end{bmatrix} & \mathbf{H}_{\{4,5\}} &= \begin{bmatrix} 0 & 4 \\ 4 & 0 \\ 2 & 2 \end{bmatrix}.
\end{aligned} \tag{91}$$

Let us see now why security is guaranteed. Intuitively, we require the key variables to fully cover the messages, which will translate to the requirement that certain matrices have full rank. Suppose we have $|\mathcal{T}| = 2$ colluding users, say $\mathcal{T} = \{4, 5\}$. Consider the security constraint (6).

$$\begin{aligned}
&I\left(W_1, W_2, W_3; X_1, X_2, X_3 | \dots \right. \\
&\qquad \left. \dots W_1 + W_2 + W_3, W_4, W_5, Z_4, Z_5\right) \\
&= H\left(X_1, X_2, X_3 | W_1 + W_2 + W_3, W_4, W_5, Z_4, Z_5\right) \\
&\quad - H\left(X_1, X_2, X_3 | W_1, W_2, W_3, W_4, W_5, Z_4, Z_5\right) \qquad (92) \\
&\overset{(90)}{\leq} 2L - H\left(\underbrace{\begin{bmatrix} \mathbf{H}_{\{1,2\}} & \mathbf{H}_{\{1,3\}} & \mathbf{0} \\ -\mathbf{H}_{\{1,2\}} & \mathbf{0} & \mathbf{H}_{\{2,3\}} \\ \mathbf{0} & -\mathbf{H}_{\{1,3\}} & -\mathbf{H}_{\{2,3\}} \end{bmatrix}}_{=\hat{\mathbf{H}}} \begin{bmatrix} S_{\{1,2\}} \\ S_{\{1,3\}} \\ S_{\{2,3\}} \end{bmatrix}\right) \qquad (93) \\
&= 2L - H\left(S_{\{1,2\}}, S_{\{1,3\}}, S_{\{2,3\}}\right) \qquad (94) \\
&= 2L - 3L_S = 2 \times 3 - 3 \times 2 = 0 \qquad (95)
\end{aligned}$$

where in (93), the first term follows from the maximum number of symbols contained in X_1, X_2, X_3 after conditioning on $X_1 + X_2 + X_3 = W_1 + W_2 + W_3$; the second term follows from the design and the independence of the keys and messages (90). To obtain (94), we require $\hat{\mathbf{H}}$ to have full rank so that the precoded keys are invertible to the original groupwise keys (i.e., the key precoding matrices are sufficiently generic). We may readily verify that $\hat{\mathbf{H}}$ has full rank for the assignment (91). Note that $\hat{\mathbf{H}}$ has the same form as (88) in the $T = 0$ case so that the assignment in (88) can also be used here (such a reduction will be used in the general proof). Therefore, we have seen how the security proof can be translated to the full rank property of the precoding matrices (see Lemma 4 for the general result). The security for other cases of colluding users can be similarly verified.

We are now ready to proceed to the general proof. As $\mathcal{R}_g^* = \emptyset$ when $G = 1$ or $G > K - T$, we only need to consider the case where $1 < G \leq K - T$.

3) General Proof for Arbitrary $K, T, 1 < G \leq K - T$: Suppose[7] $L = L_X = K!\binom{K-T}{G}M$ and $L_S = K!(K-T-1)M$ so that the desired rate is achieved. Group every M symbols from W_k together and view them as a single symbol from the extension field $\mathbb{F}_{q^M}$, i.e., $W_k \in \mathbb{F}_{q^M}^{L/M\times 1}$. Similarly, suppose $S_{\mathcal{G}} \in \mathbb{F}_{q^M}^{L_S/M\times 1}$. The messages are set as

$$X_k = W_k + \sum_{\mathcal{G}: k\in\mathcal{G}, \mathcal{G}\in\binom{[K]}{G}} \mathbf{H}_{\mathcal{G}}^k S_{\mathcal{G}}, \ \forall k \in [K] \tag{96}$$

where $\mathbf{H}_{\mathcal{G}}^k$ is an $L/M \times L_S/M$ matrix over $\mathbb{F}_{q^M}$ such that for any group, the sum of the precoded keys is zero, i.e.,

$$\sum_{k\in\mathcal{G}} \mathbf{H}_{\mathcal{G}}^k = \mathbf{0}_{L/M\times L_S/M}, \ \forall \mathcal{G} \in \binom{[K]}{G}. \tag{97}$$

[7]The input and key lengths are set to be larger than those in the examples to simplify the proof. In particular, $K!$ is introduced to allow permutation as in Section VI-B.1 to produce matrices with desired ranks. M is an integer scaling factor to work over the (larger) extension field. Finding the minimum L (sub-packetization) is left as an interesting open problem.

Collecting all messages and writing them in a matrix form, we have

$$\begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix} = \begin{bmatrix} W_1 \\ \vdots \\ W_K \end{bmatrix} + \mathbf{H} \begin{bmatrix} S_{\{1,2,\cdots,G\}} \\ \vdots \\ S_{\{K-G+1,\cdots,K\}} \end{bmatrix} \tag{98}$$

where

$$\begin{aligned} \mathbf{H} &= \left[\left(\mathbf{H}_{\mathcal{G}}^k \right)_{k\in[K],\mathcal{G}\in\binom{[K]}{G}} \right] \\ &\triangleq \begin{bmatrix} \mathbf{H}_{\{1,2,\cdots,G\}}^1 & \cdots & \mathbf{H}_{\{K-G+1,\cdots,K\}}^1 \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{\{1,2,\cdots,G\}}^K & \cdots & \mathbf{H}_{\{K-G+1,\cdots,K\}}^K \end{bmatrix} \end{aligned} \tag{99}$$

and $\mathbf{H}_{\mathcal{G}}^k = \mathbf{0}_{L/M\times L_S/M}$, $\forall k \in [K], \mathcal{G} \in \binom{[K]}{G}, k \notin \mathcal{G}$.

Correctness is straightforward, as

$$\sum_{k\in[K]} X_k = \sum_{k\in[K]} W_k + \sum_{k\in[K]} \sum_{\mathcal{G}:k\in\mathcal{G},\mathcal{G}\in\binom{[K]}{G}} \mathbf{H}_{\mathcal{G}}^k S_{\mathcal{G}} \tag{100}$$

$$= \sum_{k\in[K]} W_k + \sum_{\mathcal{G}\in\binom{[K]}{G}} \left[\left(\sum_{k\in\mathcal{G}} \mathbf{H}_{\mathcal{G}}^k \right) S_{\mathcal{G}} \right] \tag{101}$$

$$\overset{(97)}{=} \sum_{k\in[K]} W_k. \tag{102}$$

We show that security is guaranteed if some precoding matrix has certain rank, presented in the following lemma. The proof of the existence of such matrices is deferred to the next section.

Consider any colluding user set $\mathcal{T}$. Define the following submatrix of $\mathbf{H}$ (which is obtained by considering the keys only known to non-colluding users).

$$\hat{\mathbf{H}}[\mathcal{T}] \triangleq \left[\left(\mathbf{H}_{\mathcal{G}}^k \right)_{k\in[K]\backslash\mathcal{T},\mathcal{G}\in\binom{[K]\backslash\mathcal{T}}{G}} \right] \tag{103}$$

which contains $K - |\mathcal{T}|$ row blocks and $\binom{K-|\mathcal{T}|}{G}$ column blocks of $\mathbf{H}_{\mathcal{G}}^k$ terms.

Lemma 4: For any colluding user set $\mathcal{T} \subset [K], |\mathcal{T}| \leq T$, the scheme (96) satisfies the security constraint (6) if and only if $\text{rank}(\hat{\mathbf{H}}[\mathcal{T}]) = (K - |\mathcal{T}| - 1)L/M$ over $\mathbb{F}_{q^M}$.

Proof: Consider the 'if' direction. Consider the security constraint (6).

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

$$= I\left((W_k)_{k\in[K]\backslash\mathcal{T}}; (X_k)_{k\in[K]\backslash\mathcal{T}} \middle| \cdots \right.$$

$$\left. \cdots \sum_{k\in[K]\backslash\mathcal{T}} W_k, (W_k, Z_k)_{k\in\mathcal{T}} \right) \tag{104}$$

$$= H\left((X_k)_{k\in[K]\backslash\mathcal{T}} \middle| \sum_{k\in[K]\backslash\mathcal{T}} W_k, (W_k, Z_k)_{k\in\mathcal{T}} \right)$$

$$- H\left((X_k)_{k\in[K]\backslash\mathcal{T}} \middle| (W_k)_{k\in[K]}, (Z_k)_{k\in\mathcal{T}} \right) \tag{105}$$

$$= H\left((X_k)_{k\in[K]\backslash\mathcal{T}} \middle| \sum_{k\in[K]\backslash\mathcal{T}} W_k, (W_k, Z_k)_{k\in\mathcal{T}} \right)$$

$$- H\left(\left(\sum_{\mathcal{G}:k\in\mathcal{G},\mathcal{G}\in\binom{[K]\backslash\mathcal{T}}{G}} \mathbf{H}_{\mathcal{G}}^k S_{\mathcal{G}} \right)_{k\in[K]\backslash\mathcal{T}} \middle| \cdots \right.$$

$$\left. \cdots (W_k)_{k\in[K]}, (Z_k)_{k\in\mathcal{T}} \right) \tag{106}$$

$$\leq (K - |\mathcal{T}| - 1)L - H\left(\hat{\mathbf{H}}[\mathcal{T}] \left(S_{\mathcal{G}} \right)_{\mathcal{G}\in\binom{[K]\backslash\mathcal{T}}{G}} \right) \tag{107}$$

$$= (K - |\mathcal{T}| - 1)L - (K - |\mathcal{T}| - 1)L = 0 \tag{108}$$

where in (107), the first term follows from the fact that $\sum_{k\in[K]\backslash\mathcal{T}} X_k = \sum_{k\in[K]\backslash\mathcal{T}} W_k$ and uniform distribution maximizes entropy (note that the entropy is measured in q-ary units); the second term is obtained by using the definition of $\hat{\mathbf{H}}[\mathcal{T}]$ (103), the independence of the input and key variables (1), and the independence of the groupwise key variables (9). In the last step, we use the assumption that $\hat{\mathbf{H}}[\mathcal{T}]$ has rank $(K - |\mathcal{T}| - 1)L/M$ over $\mathbb{F}_{q^M}$ and $S_{\mathcal{G}}$ symbols are i.i.d. and uniform.

The 'only if' direction is obvious. For any scheme of form (96), if (104) is zero, then all inequalities above must be strictly equality, i.e., $\hat{\mathbf{H}}[\mathcal{T}]$ must have rank $(K-|\mathcal{T}|-1)L/M$. ■

4) Existence of $\mathbf{H}$*: Reduction to* $T = 0$*:* In this section, we show that there exists a matrix $\mathbf{H}$ of form (99) such that its submatrix $\hat{\mathbf{H}}[\mathcal{T}]$, specified in (103), has rank $(K - |\mathcal{T}| - 1)L/M$ over $\mathbb{F}_{q^M}$ for all possible colluding user sets $\mathcal{T}, \forall \mathcal{T} \subset [K], |\mathcal{T}| \leq T$.

We show that if for each $\mathcal{G} \in \binom{[K]}{G}$, we generate each element of any $G - 1 > 0$ matrices $\mathbf{H}_{\mathcal{G}}^k, k \in \mathcal{G}$ in (99) uniformly and i.i.d. over the extension field $\mathbb{F}_{q^M}$ (and the last matrix is set as the negative of the sum of the remaining $G - 1$ matrices, to satisfy (97)), then as[8] $M \to \infty$, the probability that $\hat{\mathbf{H}}[\mathcal{T}]$ has rank $(K-|\mathcal{T}|-1)L/M$ approaches 1 such that the existence of $\mathbf{H}$ is guaranteed. To apply the Schwartz-Zippel lemma [19], [20], [21], we need to guarantee that for each $\mathcal{T}$, there exists a realization of $\hat{\mathbf{H}}[\mathcal{T}]$ such that $\text{rank}(\hat{\mathbf{H}}[\mathcal{T}]) = (K - |\mathcal{T}| - 1)L/M$.

We are left to show that for any fixed $\mathcal{T}$, we have an $\hat{\mathbf{H}}[\mathcal{T}]$ of rank $(K-|\mathcal{T}|-1)L/M$ over $\mathbb{F}_{q^M}$. Note that $\hat{\mathbf{H}}[\mathcal{T}]$ is the same as the precoding matrix $\mathbf{H}$ in (99) when we have $K-|\mathcal{T}|$ users in $[K]\backslash\mathcal{T}$, 0 colluding users, and groupwise keys of group size G. Here from the proof of Theorem 1, we have a scalar $(L = 1)$ linear basic scheme of form (96) (where we may arbitrarily associate the coded keys to groupwise ones) and

[8]Note that here we present an existence proof over an exceedingly large block lengths based on probabilistic arguments, similar to the Shannon's original random coding proof to channel capacity. We leave the problem of finding an efficient deterministic explicit code construction as interesting future work.

then following the example in Section VI-B.1, we can permute the basic scheme to produce a length $L = K!$ scheme with symmetric groupwise keys. We can further extend the field to $\mathbb{F}_{q^M}$ by considering a block of M symbols together. Then by repeating the scheme $\binom{K-T}{G}$ times gives us the desired input length $L = K!\binom{K-T}{G}M$. Repeating and permuting the basic scheme preserves correctness and security. By the 'only if' direction of Lemma 4, such a secure scheme will produce the desired $\hat{\mathbf{H}}[\mathcal{T}]$.

VII. Conclusion

In this work, we have studied an elemental one hop information theoretic model on secure summation. Our main results include the characterization of the capacity region of secure summation with arbitrarily coded or symmetric groupwise keys and symmetric colluding user sets, and the feasibility condition of secure summation with general groupwise keys and colluding user sets.

References

[1] K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2017, pp. 1175–1191.

[2] J. H. Bell, K. A. Bonawitz, A. Gascón, T. Lepoint, and M. Raykova, "Secure single-server aggregation with (poly)logarithmic overhead," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2020, pp. 1253–1269.

[3] J. So, B. Güler, and A. S. Avestimehr, "Turbo-aggregate: Breaking the quadratic aggregation barrier in secure federated learning," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 479–489, Mar. 2021.

[4] S. Kadhe, N. Rajaraman, O. Ozan Koyluoglu, and K. Ramchandran, "FastSecAgg: Scalable secure aggregation for privacy-preserving federated learning," 2020, *arXiv:2009.11248*.

[5] 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.

[6] J. So et al., "LightSecAgg: A lightweight and versatile design for secure aggregation in federated learning," in *Proc. Mach. Learn. Syst.*, vol. 4, pp. 694–720, Apr. 2022.

[7] T. Jahani-Nezhad, M. A. Maddah-Ali, S. Li, and G. Caire, "SwiftAgg+: Achieving asymptotically optimal communication loads in secure aggregation for federated learning," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 977–989, Apr. 2023.

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

[9] R. Schlegel, S. Kumar, E. Rosnes, and A. G. I. Amat, "CodedPaddedFL and CodedSecAgg: Straggler mitigation and secure aggregation in federated learning," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2013–2027, Apr. 2023.

[10] Z. Liu, J. Guo, K.-Y. Lam, and J. Zhao, "Efficient dropout-resilient aggregation for privacy-preserving machine learning," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1839–1854, 2023.

[11] M. Hayashi and T. Koshiba, "Secure modulo zero-sum randomness as cryptographic resource," Cryptol. ePrint Arch., Tech. Paper 2018/802, 2018. [Online]. Available: https://eprint.iacr.org/2018/802

[12] Y. Zhao and H. Sun, "Expand-and-randomize: An algebraic approach to secure computation," *Entropy*, vol. 23, no. 11, p. 1461, Nov. 2021.

[13] K. Wan, H. Sun, M. Ji, and G. Caire, "On secure distributed linearly separable computation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, pp. 912–926, Mar. 2022.

[14] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[15] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. I. secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[16] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec. 2004.

[17] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals—Part I," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3973–3996, Aug. 2010.

[18] C. Chan and L. Zheng, "Mutual dependence for secret key agreement," in *Proc. 44th Annu. Conf. Inf. Sci. Syst.*, Mar. 2010, pp. 1–6.

[19] R. A. Demillo and R. J. Lipton, "A probabilistic remark on algebraic program testing," *Inf. Process. Lett.*, vol. 7, no. 4, pp. 193–195, Jun. 1978.

[20] J. T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, pp. 701–717, Oct. 1980.

[21] R. Zippel, "Probabilistic algorithms for sparse polynomials," in *Proc. Int. Symp. Symbolic Algebr. manipulation*, Berlin, Germany: Springer, 1979, pp. 216–226.

Yizhou Zhao (Member, IEEE) received the B.E. degree in communication engineering from Southwest University, Chongqing, China, in 2017, the M.S. degree in electrical and computer engineering from the University of Illinois at Chicago, USA, in 2018, and the Ph.D. degree in electrical engineering from the University of North Texas, USA, in 2023. His research interests include information theory and its applications to secure computation.

Hua Sun (Member, IEEE) received the B.E. degree in communications engineering from the Beijing University of Posts and Telecommunications, China, in 2011, and the M.S. degree in electrical and computer engineering and the Ph.D. degree in electrical engineering from the University of California, Irvine, USA, in 2013 and 2017, respectively. He is an Associate Professor with the Department of Electrical Engineering, University of North Texas, USA. His research interests include information theory and its applications to communications, privacy, security, and storage.

He was a recipient of the NSF CAREER Award in 2021, the UNT College of Engineering Junior Faculty Research Award in 2021, and the UNT College of Engineering Distinguished Faculty Fellowship in 2023. His coauthored papers received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016, the IEEE GLOBECOM Best Paper Award in 2016, and the 2020–2021 IEEE Data Storage Best Student Paper Award.