Weakly Secure Summation with Colluding Users

Zhou Li, Yizhou Zhao, Hua Sun
Department of Electrical Engineering
University of North Texas, Denton, TX 76203
Email: zhouli@my.unt.edu, yizhouzhao@my.unt.edu, hua.sun@unt.edu

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206930

***Abstract*—In secure summation, K users, each holds an input, wish to compute the sum of the inputs at a server without revealing any information about *all the inputs* even if the server may collude with *an arbitrary subset of users*. In this work, we relax the security and colluding constraints, where the set of inputs whose information is prohibited from leakage is from a predetermined collection of sets (e.g., any set of up to S inputs) and the set of colluding users is from another predetermined collection of sets (e.g., any set of up to T users). For arbitrary collection of security input sets and colluding user sets, we characterize the optimal randomness assumption, i.e., the minimum number of key bits that need to be held by the users, per input bit, for weakly secure summation to be feasible, which generally involves solving a linear program.**

I. Introduction

The focus of this work is on the information theoretic secure summation problem [1] (see Fig. 1), where User $k \in \{1, 2, \cdots, K\}$ holds an input variable W_k and an independent key variable Z_k from a finite field, and is connected to a server through a noiseless orthogonal link. From one message X_k from each user, the server shall be able to decode the sum of the inputs $W_1 + \cdots + W_K$ while obtaining no additional information about all the inputs $W_1, \cdots, W_K$.

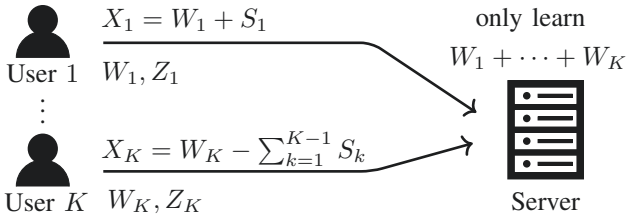

Fig. 1: The secure summation problem and an optimal protocol where $S_1, \cdots, S_{K-1}$ are uniform and independent.

An optimal secure summation protocol is plotted in Fig. 1, where the key variables (set as functions of S_k) are $(K-1)$-MDS and zero-sum, i.e., any $K-1$ variables from $Z_1, \cdots, Z_K$ are independent and uniform, and $Z_1 + \cdots + Z_K$ is 0. The optimality of the protocol is regarding both the communication cost and the randomness cost, i.e., in order to compute 1 bit of the summation securely, each user must send a message X_k of at least 1 bit to the server and the K users need to hold key variables of at least $K-1$ bits. Note that the randomness cost scales linearly with the total number of users K, which could be huge in practice. This is mainly due to the stringent security constraint, i.e., we wish to protect all K inputs. One main motivation of this work is to relax the security constraint to a weaker one (i.e., the set of inputs that need to be kept secure are some subsets of all inputs) and understand its impact on the randomness cost. Moreover, the minimum communication cost and randomness cost for secure summation remain unchanged even if user-server collusion is included [1]. In particular, no matter which set of users (big or small) may collude with the server so that the server might get some advantage in inferring information about the remaining users, the optimal protocol remains the same. The other main motivation of this work is to see if the dependence on the colluding pattern will be more explicit in the weakly secure summation problem, i.e., we wish to study the joint effect of arbitrary security and colluding patterns on the randomness consumption.

The main result of this work is a complete characterization of the minimum key size for weakly secure summation with arbitrary security and colluding patterns, i.e., arbitrary security input sets and colluding user sets. The ultimate answer generally involves two parts - one integral part that corresponds to the number of users that need to be protected under a pair of security input set and colluding user set and one possibly fractional part that corresponds to the amount of key required by remaining users and determined by a linear program.

II. Problem Statement and Definitions

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

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

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

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 $\mathbb{F}_q$.

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

User k sends to the server a message X_k, which is a function of W_k, Z_k and consists of L_X symbols from $\mathbb{F}_q$.

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

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

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

Authorized licensed use limited to: University of North Texas. Downloaded on August 31,2023 at 19:54:48 UTC from IEEE Xplore. Restrictions apply.

The security input sets are described by a monotone[1] set system $\{\mathcal{S}_1, \cdots, \mathcal{S}_M\}$ and the colluding user sets are described by another monotone set[2] system $\{\mathcal{T}_1, \cdots, \mathcal{T}_N\}$. The security constraint states that if the server colludes with users from any $\mathcal{T}_n$ set, nothing additional is revealed about the inputs from any $\mathcal{S}_m$ set (note that $\mathcal{S}_m \cap \mathcal{T}_n$ may not be empty),

$$I\left((W_k)_{k\in\mathcal{S}_m}; (X_k)_{k\in[K]} \middle| \sum_{k\in[K]} W_k, (W_k, Z_k)_{k\in\mathcal{T}_n}\right) = 0, \quad \forall m \in [M], n \in [N]. \tag{6}$$

The key *rate* R_{Z_Σ}, characterizes how many symbols the source key variable contains per input symbol, and is defined as $R_{Z_\Sigma} \triangleq L_{Z_\Sigma}/L$. The rate R_{Z_Σ} is said to be achievable if there exists a secure summation scheme, for which (5) and (6) are satisfied, and the key rate is no greater than R_{Z_Σ}. The infimum of the achievable rates R_{Z_Σ} is called the optimal key rate, denoted as $R^*_{Z_\Sigma}$.

A. Auxiliary Definitions

Definition 1 (Implicit Security Input Set $\mathcal{S}_I$): The implicit security input set is defined as

$$\mathcal{S}_I \triangleq \Big\{[K] \setminus \{\mathcal{S}_m \cup \mathcal{T}_n\} : |\mathcal{S}_m \cup \mathcal{T}_n| = K-1, \cdots \cdots \forall m \in [M], \forall n \in [N]\Big\} \setminus \{\cup_{i\in[M]}\mathcal{S}_i\}. \tag{7}$$

We will use the following example to explain the definitions.

Example 1: Consider $K = 5$, the security input sets are $(\mathcal{S}_1, \cdots, \mathcal{S}_4) = (\emptyset, \{1\}, \{2\}, \{3\})$, and the colluding user sets are $(\mathcal{T}_1, \cdots, \mathcal{T}_{14}) = (\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,5\}, \{3,4\}, \{3,5\}, \{1,3,4\}, \{2,3,5\})$.

Searching for all security input set $\mathcal{S}_m$ and colluding user set $\mathcal{T}_n$ whose union has cardinality $K-1=4$, we have $|\mathcal{S}_2 \cup \mathcal{T}_{14}| = |\{1\} \cup \{2,3,5\}| = 4$ and $|\mathcal{S}_3 \cup \mathcal{T}_{13}| = |\{2\} \cup \{1,3,4\}| = 4$, so $\mathcal{S}_I = \{4, 5\}$ for Example 1.

Definition 2 (Total Security Input Set $\overline{\mathcal{S}}$): The union of explicit and implicit security input sets is defined as the total security input set, $\overline{\mathcal{S}} \triangleq \cup_{m\in[M]}\mathcal{S}_m \cup \mathcal{S}_I$.

For Example 1, we have $\overline{\mathcal{S}} = \{1, 2, 3, 4, 5\}$.

Definition 3 (Intersection of $\mathcal{S}_m \cup \mathcal{T}_n$ and $\overline{\mathcal{S}}$, $\mathcal{A}_{m,n}$): For each pair of security input set $\mathcal{S}_m$ and colluding user set $\mathcal{T}_n$, its overlap with the total security input set $\overline{\mathcal{S}}$ is denoted as

$$\mathcal{A}_{m,n} \triangleq (\mathcal{S}_m \cup \mathcal{T}_n) \cap \overline{\mathcal{S}} \tag{8}$$

and its maximum cardinality is denoted as

$$a^* \triangleq \max_{m\in[M], n\in[N]} |\mathcal{A}_{m,n}|. \tag{9}$$

For Example 1, $\mathcal{A}_{2,14} = (\{1\} \cup \{2,3,5\}) \cap \{1,2,3,4,5\} = \{1,2,3,5\}$, and $\mathcal{A}_{3,13} = (\{2\} \cup \{1,3,4\}) \cap \{1,2,3,4,5\} = \{1,2,3,4\}$. Further, $a^* = 4$.

[1]A set system is called monotone if a set belongs to the system, then its subset also belongs to the system.

[2]Without loss of generality, assume $\cup_m \mathcal{S}_m \neq \emptyset$ (the security constraints are not empty) and $|\mathcal{T}_n| \leq K-2$ as otherwise there is nothing to hide.

Definition 4 (Union of Maximum $\mathcal{A}_{m,n}$): Find all $\mathcal{A}_{m,n}$ sets with maximum cardinality and denote the union of the corresponding $\mathcal{S}_m, \mathcal{T}_n$ sets as $\mathcal{Q}$.

$$\mathcal{Q} \triangleq \cup_{m,n:|\mathcal{A}_{m,n}|=a^*} \mathcal{S}_m \cup \mathcal{T}_n. \tag{10}$$

For Example 1, $\mathcal{A}_{2,14}, \mathcal{A}_{3,13}$ are all $\mathcal{A}_{m,n}$ sets with the maximum cardinality, so $\mathcal{Q} = \mathcal{S}_2 \cup \mathcal{T}_{14} \cup \mathcal{S}_3 \cup \mathcal{T}_{13} = \{1,2,3,4,5\}$.

III. Result

Theorem 1: For secure summation with $K \geq 2$ users, security input sets $(\mathcal{S}_m)_{m\in[M]}$, and colluding user sets $(\mathcal{T}_n)_{n\in[N]}$, the optimal key rate $R^*_{Z_\Sigma}$ is

$$R^*_{Z_\Sigma} = \begin{cases} a^* + b^* & \text{if } a^* \leq K-1, \\ & a^* = |\overline{\mathcal{S}}|, |\mathcal{Q}| = K \\ \min(a^*, K-1) & \text{otherwise} \end{cases} \tag{11}$$

where b^* is the optimal value of the following linear program[3],

$$\min \max_{m,n:|\mathcal{A}_{m,n}|=a^*} \sum_{k\in\mathcal{T}_n\setminus\overline{\mathcal{S}}} b_k$$
$$\text{subject to} \sum_{k\in[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)} b_k \geq 1, \forall m, n \text{ s.t. } |\mathcal{A}_{m,n}| = a^*,$$
$$b_k \geq 0, \forall k \in [K] \setminus \overline{\mathcal{S}}. \tag{12}$$

IV. Converse Proof of Theorem 1

We start from the 'otherwise' case and show that $R_{Z_\Sigma} \geq \min(a^*, K-1)$. Let's use Example 1 to illustrate the idea.

A. Proof of Example 1

The converse proof is based on showing that for any $\mathcal{A}_{m,n}$, we have $H\big((Z_k)_{k\in\mathcal{A}_{m,n}}\big) \geq |\mathcal{A}_{m,n}|L$.

For Example 1, let's take $\mathcal{A}_{3,13} = \{1,2,3,4\}$ as an example, where $\{4\}$ comes from the implicit security input set $\mathcal{S}_I$ and $\{1,2,3\}$ comes from the explicit security input set $\cup_m \mathcal{S}_m$. When expanding $H\big((Z_k)_{k\in\mathcal{A}_{m,n}}\big)$, we first consider the term from the implicit set and then consider the term from the explicit set (conditioned on the implicit set), i.e., $H(Z_1, Z_2, Z_3, Z_4) = H(Z_4) + H(Z_1, Z_2, Z_3|Z_4)$. Next, we show that $H(Z_4) \geq L$ and $H(Z_1, Z_2, Z_3|Z_4) \geq 3L$.

First, consider $H(Z_4) \geq L$. The intuition is that $\{4\} = [K] \setminus (\mathcal{S}_2 \cup \mathcal{T}_{14}) = [5] \setminus (\{1\} \cup \{2,3,5\})$ belongs to the implicit security input set, i.e., when the server colludes with users in $\mathcal{T}_{14}$, the sum $\sum_{k\in[K]} W_k$ can be decoded and nothing is revealed about $(W_k)_{k\in\mathcal{S}_2}$. Complementarily, nothing shall be revealed about $(W_k)_{k\in[K]\setminus(\mathcal{S}_2\cup\mathcal{T}_{14})} = W_4$. Expressing this idea in entropy terms, we have

$$\begin{aligned} & H(Z_4) \geq H(Z_4|Z_2, Z_3, Z_5) \geq I(Z_4; Z_1|Z_2, Z_3, Z_5) \\ \overset{(1)}{=}\ & I(Z_4, W_4; Z_1, W_1|Z_2, Z_3, Z_5, W_2, W_3, W_5) && (13) \\ \overset{(4)}{\geq}\ & I(X_4, W_4; X_1, W_1|Z_2, Z_3, Z_5, W_2, W_3, W_5) && (14) \\ =\ & H(W_1|Z_2, Z_3, Z_5, W_2, W_3, W_5, X_1) \\ & - H(W_1|Z_2, Z_3, Z_5, W_2, W_3, W_5, X_1, X_4, W_4) && (15) \end{aligned}$$

[3]Note that for the 'if' case, $|\mathcal{S}_m \cup \mathcal{T}_n| \leq K-1$ (because otherwise $a^* = K$) and $\overline{\mathcal{S}} \subset (\mathcal{S}_m \cup \mathcal{T}_n)$ (so $b_k, k \in [K] \setminus \overline{\mathcal{S}}$ are all the variables.)

Authorized licensed use limited to: University of North Texas. Downloaded on August 31,2023 at 19:54:48 UTC from IEEE Xplore. Restrictions apply.

$$\overset{(4)(5)}{\geq} H\Big(W_1\Big|Z_2,Z_3,Z_5,W_2,W_3,W_5,(X_k)_{k\in[5]},\sum_{k\in[5]}W_k\Big) - H\Big(W_1\Big|Z_2,Z_3,Z_5,W_2,W_3,W_5,X_2,X_3,X_5,\cdots \cdots X_1,X_4,W_4,\sum_{k\in[5]}W_k\Big) \quad (16)$$

$$\overset{(6)}{\geq} H\Big(W_1\Big|Z_2,Z_3,Z_5,W_2,W_3,W_5,\sum_{k\in[5]}W_k\Big) - H(W_1|W_1) \overset{(1)(2)}{=} L-0=L \quad (17)$$

where in (17), the first term follows from (6) with $\mathcal{S}_2=\{1\}$ and $\mathcal{T}_{14}=\{2,3,5\}$; the second term follows from only keeping W_1 (obtained from $\sum_k W_k$ and W_2,W_3,W_5,W_4).

Second, consider $H(Z_1,Z_2,Z_3|Z_4)\geq 3L$. Note that $\mathcal{S}_3\cup\mathcal{T}_{13}=\{2\}\cup\{1,3,4\}=\{1,2,3,4\}$ so that Z_4 may appear in the conditioning term and $(\mathcal{S}_3\cup\mathcal{T}_{13})\cap\cup_{m\in[4]}\mathcal{S}_m=\{1,2,3\}$ so that we want to show that Z_1,Z_2,Z_3 must each contribute L independent amount of information.

$$\begin{aligned}
& H(Z_1,Z_2,Z_3|Z_4)\\
\geq\ & H(Z_1,Z_2,Z_3|Z_4,W_1,W_2,W_3,W_4) && (18)\\
\geq\ & I(Z_1,Z_2,Z_3;X_1,X_2,X_3|Z_4,W_1,W_2,W_3,W_4) && (19)\\
\overset{(4)}{=}\ & H(X_1,X_2,X_3|Z_4,W_1,W_2,W_3,W_4) && (20)\\
=\ & H(X_1,X_2,X_3|W_4,Z_4)\\
& - I(X_1,X_2,X_3;W_1,W_2,W_3|W_4,Z_4) && (21)\\
\geq\ & 3L - I(X_1,X_2,X_3;W_1|W_4,Z_4)\\
& - I(X_1,X_2,X_3;W_3|W_4,Z_4,W_1)\\
& - I(X_1,X_2,X_3;W_2|W_4,Z_4,W_1,W_3) && (22)\\
\geq\ & 3L - I\Big(X_1,X_2,X_3,\sum_{k\in[5]}W_k;W_1\Big|W_4,Z_4\Big)\\
& - I\Big(X_1,X_2,X_3,\sum_{k\in[5]}W_k,Z_1;W_3\Big|W_4,Z_4,W_1\Big)\\
& - I\Big(X_1,X_2,X_3,\sum_{k\in[5]}W_k,Z_1,Z_3;W_2\Big|\cdots\\
& \cdots W_4,Z_4,W_1,W_3\Big) && (23)\\
\overset{(1)}{=}\ & 3L - I\Big(X_1,X_2,X_3;W_1\Big|\sum_{k\in[5]}W_k,W_4,Z_4\Big)\\
& - I\Big(X_1,X_2,X_3;W_3\Big|\sum_{k\in[5]}W_k,W_4,Z_4,W_1,Z_1\Big)\\
& - I\Big(X_1,X_2,X_3;W_2\Big|\cdots\\
& \cdots\sum_{k\in[5]}W_k,W_4,Z_4,W_1,Z_1,W_3,Z_3\Big)\overset{(6)}{=}3L && (24)
\end{aligned}$$

where in (22), $H(X_1,X_2,X_3|W_4,Z_4)\geq 3L$ will be proved in Lemma 1 and the remaining terms follow from applying security constraints for various security input set and colluding user set. In (24), the second term is zero due to (6) with $\mathcal{S}_2=\{1\}$ and $\mathcal{T}_5=\{4\}$, the third term is zero due to (6) with $\mathcal{S}_4=\{3\}$ and $\mathcal{T}_8=\{1,4\}$ and the fourth term is zero due to (6) with $\mathcal{S}_3=\{2\}$ and $\mathcal{T}_{13}=\{1,3,4\}$. Note that the order of chain-rule expansion is carefully chosen, where W_2 is considered last as it belongs to $\mathcal{S}_3$ while the other terms W_1,W_3 are considered first as they belong to $\mathcal{T}_{13}$ and the set systems are monotone (remember that the term of consideration $H(Z_1,Z_2,Z_3,Z_4)$ comes from $\mathcal{A}_{3,13}$).

B. Proof of $R_{Z_\Sigma}\geq\min(a^*,K-1)$

We are now ready to generalize the above proof to all parameter settings. The ideas are captured by four lemmas, whose proofs are deferred to [2]. First, each X_k must contain L symbols even if all other inputs are known.

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

$$H\big(X_u\big|(W_k,Z_k)_{k\in[K]\setminus\{u\}}\big)\geq L. \quad (25)$$

Next, the keys used by users outside $\mathcal{S}_m\cup\mathcal{T}_n$ should not be less than L symbols, conditioned on the colluding information.

Lemma 2: For any $\mathcal{S}_m,\mathcal{T}_n,m\in[M],n\in[N]$ such that $\mathcal{S}_m\cap\mathcal{T}_n=\emptyset$ and $|\mathcal{S}_m\cup\mathcal{T}_n|\leq K-1$, we have

$$H\big((Z_k)_{k\in[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)}\big|(Z_k)_{k\in\mathcal{T}_n}\big)\geq L. \quad (26)$$

Consider $|\mathcal{S}_m\cup\mathcal{T}_n|=K-1$ for Lemma 2.

Corollary 1: For any $\mathcal{S}_m,\mathcal{T}_n$, such that $\mathcal{S}_m\cap\mathcal{T}_n=\emptyset$ and $|\mathcal{S}_m\cup\mathcal{T}_n|=K-1$, denote $u=[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)$ and we have

$$H\big(Z_u\big|(Z_k)_{k\in\mathcal{T}_n}\big)\geq L. \quad (27)$$

Now for any $\mathcal{S}_m,\mathcal{T}_n$, the keys used by users in $(\mathcal{S}_m\cup\mathcal{T}_n)\cap(\cup_{i\in[M]}\mathcal{S}_i)$ must be at least its cardinality times L.

Lemma 3: For any $\mathcal{S}_m,\mathcal{T}_n,m\in[M],n\in[N]$ such that $\mathcal{S}_m\cap\mathcal{T}_n=\emptyset$ and $|\mathcal{S}_m\cup\mathcal{T}_n|\leq K-1$, we have

$$\begin{aligned}
& H\big((Z_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)\cap(\cup_{i\in[M]}\mathcal{S}_i)}\big|(Z_k)_{k\in\mathcal{T}_n\setminus(\cup_{i\in[M]}\mathcal{S}_i)}\big)\\
\geq\ & |(\mathcal{S}_m\cup\mathcal{T}_n)\cap(\cup_{i\in[M]}\mathcal{S}_i)|L. && (28)
\end{aligned}$$

Generalize the above lemma to also include $\mathcal{S}_I$.

Lemma 4: For any $\mathcal{S}_m,\mathcal{T}_n,m\in[M],n\in[N]$ such that $\mathcal{S}_m\cap\mathcal{T}_n=\emptyset$ and $|\mathcal{S}_m\cup\mathcal{T}_n|\leq K-1$, we have

$$\begin{aligned}
& H\big((Z_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)\cap\overline{\mathcal{S}}}\big|(Z_k)_{k\in\mathcal{T}_n\setminus\overline{\mathcal{S}}}\big)\\
\geq\ & \big|(\mathcal{S}_m\cup\mathcal{T}_n)\cap\overline{\mathcal{S}}\big|L. && (29)
\end{aligned}$$

The proof of $R_{Z_\Sigma}\geq\min(a^*,K-1)$ follows immediately from Lemma 4 as a^* is defined as the maximum cardinality of set on the RHS of (29) (refer to (8), (9)). Note that when $a^*=K$, there exist $\mathcal{S}_m,\mathcal{T}_n$ so that $|\mathcal{S}_m\cup\mathcal{T}_n|=K-1$ and then we can apply Lemma 4.

C. Proof of Example 2

Next we consider the 'if' case of Theorem 1, where we need one more step to further tighten the bound $R_{Z_\Sigma}\geq a^*$ to include an additional term b^*. To appreciate the idea in a simpler setting, let's again start with an example.

Example 2: Consider $K=5$, $(\mathcal{S}_1,\mathcal{S}_2,\mathcal{S}_3)=(\emptyset,\{1\},\{2\})$, and $(\mathcal{T}_1,\cdots,\mathcal{T}_9)=(\emptyset,\{1\},\cdots,\{5\},\{1,3\},\{2,4\},\{2,5\})$.

For Example 2, $\mathcal{A}_{2,3}=\mathcal{A}_{3,2}=\mathcal{A}_{2,8}=\mathcal{A}_{2,9}=\mathcal{A}_{3,7}=\overline{\mathcal{S}}=\{1,2\}$, and $a^*=\big|\overline{\mathcal{S}}\big|=2\leq K-1=4$. Then $\mathcal{Q}=\cup_{m,n:|\mathcal{A}_{m,n}|=a^*}\mathcal{S}_m\cup\mathcal{T}_n=\{1,2,3,4,5\}$ and $|\mathcal{Q}|=5=K$. So we are in the 'if' case.

Authorized licensed use limited to: University of North Texas. Downloaded on August 31,2023 at 19:54:48 UTC from IEEE Xplore. Restrictions apply.

Consider all $\mathcal{A}_{m,n}$ sets so that $|\mathcal{A}_{m,n}| = a^*$. For Example 2, we have 5 such sets $\mathcal{A}_{2,3}$, $\mathcal{A}_{3,2}$, $\mathcal{A}_{2,8}$, $\mathcal{A}_{2,9}$, $\mathcal{A}_{3,7}$. Note that $\mathcal{S}_m \cup \mathcal{T}_n = \mathcal{A}_{m,n} \cup (\mathcal{T}_n \setminus \overline{\mathcal{S}})$, and we consider the key variables Z_k in the set $\mathcal{S}_m \cup \mathcal{T}_n$ and split them to $\mathcal{A}_{m,n}$ (treated by Lemma 4 and corresponds to a^*) and $\mathcal{T}_n \setminus \overline{\mathcal{S}}$ (the new part treated by a linear program and corresponds to b^*).

$$\begin{aligned} H(Z_\Sigma) &\geq \max\big(H\big((Z_k)_{k\in\mathcal{S}_2\cup\mathcal{T}_3}\big), H\big((Z_k)_{k\in\mathcal{S}_3\cup\mathcal{T}_2}\big), \\ &\quad H\big((Z_k)_{k\in\mathcal{S}_2\cup\mathcal{T}_8}\big), H\big((Z_k)_{k\in\mathcal{S}_2\cup\mathcal{T}_9}\big), H\big((Z_k)_{k\in\mathcal{S}_3\cup\mathcal{T}_7}\big)\big) \\ &= \max\big(H(Z_4) + H(Z_1, Z_2|Z_4), H(Z_5) + H(Z_1, Z_2|Z_5), \\ &\quad H(Z_3) + H(Z_1, Z_2|Z_3)\big) \qquad (30) \\ &\overset{(29)}{\geq} \max\big(H(Z_4), H(Z_5), H(Z_3)\big) + a^*L \qquad (31) \end{aligned}$$

where in (30), we split the Z_k term in set $\mathcal{S}_m \cup \mathcal{T}_n$ to that in $\mathcal{T}_n \setminus \overline{\mathcal{S}}$ and $\mathcal{A}_{m,n}$; in (31), we use Lemma 4 to bound the $\mathcal{A}_{m,n}$ term conditioned on $\mathcal{T}_n \setminus \overline{\mathcal{S}}$. Next, we bound the term $\max(H(Z_4), H(Z_5), H(Z_3))$, where it turns out that the only constraints required are from Lemma 2. From (26), we have

$$\begin{aligned} \mathcal{S}_2, \mathcal{T}_8 : &\quad H(Z_3, Z_5|Z_2, Z_4) \geq L, \\ \mathcal{S}_2, \mathcal{T}_9 : &\quad H(Z_3, Z_4|Z_2, Z_5) \geq L, \\ \mathcal{S}_3, \mathcal{T}_7 : &\quad H(Z_4, Z_5|Z_1, Z_3) \geq L. \end{aligned} \qquad (32)$$

To bound (31) with constraints in (32), we resort to a linear program, with conditional entropy terms as the variables that are consistent with chain-rule expansion. In particular, set $H(Z_3) = b_3 L, H(Z_4|Z_3) = b_4 L, H(Z_5|Z_3, Z_4) = b_5 L$, then

$$R_{Z_\Sigma} \quad \geq \quad a^* + \min\max(b_3, b_4, b_5) \qquad (33)$$

where $\min$ is over the following linear constraints,

$$\begin{aligned} &b_3 + b_5 \geq \big(H(Z_3|Z_2, Z_4) + H(Z_5|Z_2, Z_3, Z_4)\big)/L \geq 1, \\ &b_3 + b_4 \geq \big(H(Z_3|Z_2, Z_5) + H(Z_4|Z_2, Z_3, Z_5)\big)/L \geq 1, \\ &b_4 + b_5 \geq \big(H(Z_4|Z_1, Z_3) + H(Z_5|Z_1, Z_3, Z_4)\big)/L \geq 1, \\ &b_3, b_4, b_5 \geq 0. \end{aligned} \qquad (34)$$

Intuitively, the correlation/conflict between $H(Z_3)$, $H(Z_4)$, $H(Z_5)$ is captured through (32) and we wish to find the tightest bound subject to (32) (interestingly and somewhat surprisingly, this bound turns out to be tight). Therefore we have transformed the R_{Z_Σ} converse to a linear program on non-negative variables b_3, b_4, b_5, where each $\mathcal{A}_{m,n}$ set with maximum cardinality contributes one linear constraint (some redundant ones are removed in (34)). Now defining $b^* = \min\max(b_3, b_4, b_5)$ as the optimal value of the linear program subject to constraints (34), we have obtained the desired converse $R_{Z_\Sigma} \geq a^* + b^*$.

D. Proof of $R_{Z_\Sigma} \geq a^* + b^*$

Building upon the insights from Example 2, we present the general proof of $R_{Z_\Sigma} \geq a^* + b^*$ when the 'if' condition holds, i.e., $a^* \leq K - 1$, $a^* = |\overline{\mathcal{S}}|$, and $|\mathcal{Q}| = K$.

Note that $a^* = |\overline{\mathcal{S}}|$ so that each $\mathcal{A}_{m,n}$ such that $|\mathcal{A}_{m,n}| = a^*$ must satisfy $\mathcal{A}_{m,n} = \overline{\mathcal{S}}$. Consider all such $\mathcal{A}_{m,n}$ sets (without loss of generality, we assume for each such $\mathcal{A}_{m,n}$, $\mathcal{S}_m \cap \mathcal{T}_n = \emptyset$ and $|\mathcal{S}_m \cup \mathcal{T}_n| \leq K - 1$ as the set systems are monotone and this will not change the linear program (12), i.e., the dropped ones are redundant). Following a similar decomposition as that of Example 2, we have

$$\begin{aligned} H(Z_\Sigma) &\overset{(3)}{\geq} \max_{m,n:|\mathcal{A}_{m,n}|=a^*} H\big((Z_k)_{k\in\mathcal{S}_m\cap\mathcal{T}_n}\big) \qquad (35) \\ &\geq \max_{m,n:|\mathcal{A}_{m,n}|=a^*} \Big(H\big((Z_k)_{k\in\mathcal{T}_n\setminus\overline{\mathcal{S}}}\big) \\ &\quad + H\big((Z_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)\cap\overline{\mathcal{S}}}\big|(Z_k)_{k\in\mathcal{T}_n\setminus\overline{\mathcal{S}}}\big)\Big) \qquad (36) \\ &\overset{(29)}{\geq} \max_{m,n:|\mathcal{A}_{m,n}|=a^*} H\big((Z_k)_{k\in\mathcal{T}_n\setminus\overline{\mathcal{S}}}\big) + a^*L \qquad (37) \end{aligned}$$

subject to the following constraints by Lemma 2,

$$\begin{aligned} &\forall m, n \text{ where } |\mathcal{A}_{m,n}| = a^* : \\ &H\big((Z_k)_{k\in[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)}\big|(Z_k)_{k\in\mathcal{T}_n}\big) \overset{(26)}{\geq} L. \end{aligned} \qquad (38)$$

Next, following the steps of the proof of Example 2, we translate the inequality (37) and the constraints (38) to a linear program in b_k variables, defined as follows.

$$\forall k \in [K] \setminus \overline{\mathcal{S}}, \quad b_k \triangleq H\Big(Z_k\big|(Z_l)_{l\in[K]\setminus\overline{\mathcal{S}}, l<k}\Big)/L. \qquad (39)$$

Then normalizing (37) by L on both hand sides and expanding the entropy term in (37) and (38) by chain-rule with lexicographic order, we have

$$R_{Z_\Sigma} \geq a^* + \min \max_{m,n:|\mathcal{A}_{m,n}|=a^*} \left(\sum_{k\in\mathcal{T}_n\setminus\overline{\mathcal{S}}} b_k\right) \qquad (40)$$

$$\begin{aligned} \text{subject to} \sum_{k\in[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)} b_k \geq 1, \forall m, n \text{ s.t. } |\mathcal{A}_{m,n}| = a^*, \\ b_k \geq 0, k \in [K] \setminus \overline{\mathcal{S}}. \end{aligned} \qquad (41)$$

Note that $\mathcal{A}_{m,n} = \overline{\mathcal{S}} \subset (\mathcal{S}_m \cup \mathcal{T}_n)$, so the chain-rule expansion above can be bounded by b_k terms. According to the linear program (12) whose optimal value is defined as b^*, we have obtained the desired converse bound $R_{Z_\Sigma} \geq a^* + b^*$.

V. Achievability Proof of Theorem 1

We similarly start from the simpler 'otherwise' case. Henceforth, we assume $a^* \leq K - 1$ because otherwise $a^* = K$ we may apply the scheme in Fig. 1 which achieves $R_{Z_\Sigma} = K - 1$ and has been proved to be correct and secure for $\mathcal{S}_m = [K]$ (so for any other $\mathcal{S}_m$) and any $\mathcal{T}_n$ in Theorem 1 of [1]. The achievable scheme of $R_{Z_\Sigma} = a^*$ for the remaining settings of the 'otherwise' case is fairly straightforward, which contains two cases and is deferred to Section 5 of [2]. We are left with the 'if' case.

A. Achievable Scheme of $R_{Z_\Sigma} = a^* + b^*$ for 'If' Case

Consider the 'if' case, where every user will be assigned some key variables, the amount of which is according to the optimal solution of the linear program (12). Denote the $b_k, k \in [K] \setminus \overline{\mathcal{S}}$ values that attain the optimal value b^* of (12) are

$$b_k = b_k^* = \frac{p_k}{\overline{q}}, \forall k \in [K] \setminus \overline{\mathcal{S}} \qquad (42)$$

Authorized licensed use limited to: University of North Texas. Downloaded on August 31,2023 at 19:54:48 UTC from IEEE Xplore. Restrictions apply.

where the linear program has rational coefficients so that the optimal solution is also rational, i.e., $p_k, \overline{q}$ are integers (and non-negative). As a result,

$$\sum_{k\in[K]\setminus\overline{\mathcal{S}}} b_k^* = \frac{\sum_{k\in[K]\setminus\overline{\mathcal{S}}} p_k}{\overline{q}} \triangleq \frac{\overline{p}}{\overline{q}}. \tag{43}$$

Pick B so that $q^B > (a^*+b^*)\overline{q}\binom{K\overline{q}}{(a^*+b^*)\overline{q}}$ and operate over $\mathbb{F}_{q^B}$. Consider $\overline{p}+(a^*-1)\overline{q}$ i.i.d. uniform variables, $\mathbf{s} = (S_1; \cdots; S_{\overline{p}+(a^*-1)\overline{q}}) \in \mathbb{F}_{q^B}^{(\overline{p}+(a^*-1)\overline{q})\times 1}$ and set the key variables as (suppose $\overline{\mathcal{S}} = \{k_1, \cdots, k_{|\overline{\mathcal{S}}|}\}$)

$$\begin{aligned} Z_k &= \mathbf{F}_k \times \mathbf{G}_k \times \mathbf{s}, \forall k \in [K]\setminus\overline{\mathcal{S}} \\ Z_k &= \mathbf{H}_k \times \mathbf{s}, \forall k \in \overline{\mathcal{S}} \end{aligned} \tag{44}$$

where each element of $\mathbf{F}_i \in \mathbb{F}_{q^B}^{\overline{q}\times p_k}$, $\mathbf{G}_i \in \mathbb{F}_{q^B}^{p_k\times(\overline{p}+(a^*-1)\overline{q})}$, $i \in [K]\setminus\overline{\mathcal{S}}$, $\mathbf{H}_j \in \mathbb{F}_{q^B}^{\overline{q}\times(\overline{p}+(a^*-1)\overline{q})}, j \in \{k_1, \cdots, k_{|\overline{\mathcal{S}}|-1}\}$ are drawn uniformly and i.i.d. from $\mathbb{F}_{q^B}$ and

$$\mathbf{H}_{k_{|\overline{\mathcal{S}}|}} = -\left(\sum_{i\in[K]\setminus\overline{\mathcal{S}}} \mathbf{F}_i \times \mathbf{G}_i + \sum_{j\in\overline{\mathcal{S}}\setminus\{k_{|\overline{\mathcal{S}}|}\}} \mathbf{H}_j\right) \tag{45}$$

$$\Rightarrow \quad \sum_{k\in[K]} Z_k \overset{(44)(45)}{=} 0. \tag{46}$$

Finally, set $L = B\overline{q}$, i.e., $W_k = (W_{k,1}; \cdots; W_{k,\overline{q}}) \in \mathbb{F}_{q^B}^{\overline{q}\times 1}$ and the sent messages as

$$X_k = W_k + Z_k, \forall k \in [K]. \tag{47}$$

Correctness is guaranteed by taking $\sum_{k\in[K]} X_k$ and (46). Note that the key rate achieved is $R_{Z_\Sigma} = L_{Z_\Sigma}/L = (\overline{p}+(a^*-1)\overline{q})/\overline{q} \overset{(43)}{=} a^* + \sum_{k\in[K]\setminus\overline{\mathcal{S}}} b_k^* - 1 \overset{(48)}{=} a^* + b^*$, where the last step is based on a crucial property of the linear program (12), stated below and proved in Lemma 5 of [2].

Lemma 5: For the linear program (12), its optimal value b^* and optimal solution b_k^* satisfy

$$b^* = \sum_{k\in[K]\setminus\overline{\mathcal{S}}} b_k^* - 1. \tag{48}$$

B. Proof of Security

Consider any set $\mathcal{S}_m, \mathcal{T}_n, m \in [M], n \in [N]$ so that[4] $|\mathcal{S}_m \cup \mathcal{T}_n| \le K-1$. We show that the security constraint (6) is satisfied for the achievable schemes under all cases.

$$\begin{aligned} & I\Big((W_k)_{k\in\mathcal{S}_m}; (X_k)_{k\in[K]}\Big|\sum_{k\in[K]} W_k, (W_k, Z_k)_{k\in\mathcal{T}_n}\Big) \\ = \; & H\Big((W_k+Z_k)_{k\in[K]}\Big|\sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in\mathcal{T}_n}\Big) \\ & - H\Big((W_k+Z_k)_{k\in[K]}\Big|\cdots \\ & \cdots \sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in\mathcal{T}_n}, (W_k)_{k\in\mathcal{S}_m}\Big) \qquad (49) \\ = \; & H\Big((W_k+Z_k)_{k\in[K]}\Big|\sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in\mathcal{T}_n}\Big) \end{aligned}$$

$$\begin{aligned} & - H\Big((W_k+Z_k)_{k\in[K]\setminus\mathcal{T}_n}\Big|\cdots \\ & \cdots \sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in\mathcal{T}_n}, (W_k)_{k\in\mathcal{S}_m}\Big) \qquad (50) \\ = \; & H\Big((W_k+Z_k)_{k\in[K]}\Big|\sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in\mathcal{T}_n}\Big) \\ & - H\Big((W_k+Z_k)_{k\in\mathcal{S}_m\setminus\mathcal{T}_n}\Big|\cdots \\ & \cdots \sum_{k\in[K]} W_k, (W_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)}, (Z_k)_{k\in\mathcal{T}_n}\Big) \\ & - H\Big((W_k+Z_k)_{k\in[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)}\Big|\cdots \\ & \cdots \sum_{k\in[K]} W_k, (W_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)}, (Z_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)}\Big) \qquad (51) \\ \overset{(1)}{=} \; & H\Big((W_k+Z_k)_{k\in[K]}\Big|\sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in\mathcal{T}_n}\Big) \\ & - H\Big((Z_k)_{k\in\mathcal{S}_m\setminus\mathcal{T}_n}\Big|(Z_k)_{k\in\mathcal{T}_n}\Big) \\ & - H\Big((W_k+Z_k)_{k\in[K]\setminus(\mathcal{S}_m\cup\mathcal{T}_n)}\Big|\cdots \\ & \cdots \sum_{k\in[K]} W_k, (W_k,Z_k)_{k\in(\mathcal{S}_m\cup\mathcal{T}_n)}\Big) \qquad (52) \\ \le \; & |\mathcal{S}_m\setminus\mathcal{T}_n|L - |\mathcal{S}_m\setminus\mathcal{T}_n|L = 0 \qquad (53) \end{aligned}$$

where in (52), the difference of the first term and the third term is no greater than $|\mathcal{S}_m\setminus\mathcal{T}_n|L$, and the second term is also $|\mathcal{S}_m\setminus\mathcal{T}_n|L$, both proved in [2] (refer to Section 5.4).

VI. Discussion

In this work, we have characterized the fundamental limits of weakly secure summation with arbitrary security constraints (where the weak security notion is similar to and a generalization of that considered in the network coding context [3]–[5]), and arbitrary colluding constraints (similar to those in private information retrieval [6], [7]). As the security and colluding constraints can be arbitrarily heterogeneous, it turns out that interestingly, their interaction can be captured by a linear program with a number of linear constraints that on the one hand, impose the security constraint for each security input and colluding user set and on the other hand, attempt to minimize the key size (the max objective function in (12) can be transformed to constraints on the additional key consumption). The resolving of such tension gives rise to matching converse claim and achievability argument (connected by the crucial algebraic property of the linear program in Lemma 5) so that the exact information theoretic answer is obtained.

Going forward, secure summation is an information theoretic primitive whose model can be further enriched to catch new requirements in federated learning, e.g., user dropout [8]–[15], user selection [16]–[19], groupwise keys [1], [14] etc. Considerations of weak security constraints in these settings are promising directions for novel insights.

Acknowledgement

This work is supported in part by NSF under Grant CCF-2007108 and Grant CCF-2045656.

[4]Recall that $a^* \le K-1$, then $|\mathcal{S}_m \cup \mathcal{T}_n|$ cannot be $[K]$.

Authorized licensed use limited to: University of North Texas. Downloaded on August 31,2023 at 19:54:48 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Y. Zhao and H. Sun, "Secure Summation: Capacity Region, Groupwise Key, and Feasibility," *arXiv preprint arXiv:2205.08458*, 2022.

[2] Z. Li, Y. Zhao, and H. Sun, "Weakly Secure Summation with Colluding Users," *arXiv preprint arXiv:2304.09771*, 2023.

[3] K. Bhattad and K. R. Narayanan, "Weakly Secure Network Coding," *Proceedings of NetCod*, 2005.

[4] D. Silva and F. R. Kschischang, "Universal Weakly Secure Network Coding," in *2009 IEEE Information Theory Workshop on Networking and Information Theory*. IEEE, 2009, pp. 281–285.

[5] M. Yan, A. Sprintson, and I. Zelenko, "Weakly Secure Data Exchange with Generalized Reed Solomon Codes," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 1366–1370.

[6] X. Yao, N. Liu, and W. Kang, "The Capacity of Private Information Retrieval under Arbitrary Collusion Patterns for Replicated Databases," *IEEE Transactions on Information Theory*, vol. 67, no. 10, pp. 6841–6855, 2021.

[7] J. Cheng, N. Liu, W. Kang, and Y. Li, "The Capacity of Symmetric Private Information Retrieval under Arbitrary Collusion and Eavesdropping Patterns," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3037–3050, 2022.

[8] 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.

[9] J. So, B. Güler, and A. S. Avestimehr, "Turbo-Aggregate: Breaking the Quadratic Aggregation Barrier in Secure Federated Learning," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 479–489, 2021.

[10] S. Kadhe, N. Rajaraman, O. O. Koyluoglu, and K. Ramchandran, "FastSecAgg: Scalable Secure Aggregation for Privacy-Preserving Federated Learning," *arXiv preprint arXiv:2009.11248*, 2020.

[11] 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.

[12] J. So, C. J. Nolet, C.-S. Yang, S. Li, Q. Yu, R. E Ali, B. Guler, and S. Avestimehr, "LightSecAgg: a Lightweight and Versatile Design for Secure Aggregation in Federated Learning," *Proceedings of Machine Learning and Systems*, vol. 4, pp. 694–720, 2022.

[13] T. Jahani-Nezhad, M. A. Maddah-Ali, S. Li, and G. Caire, "SwiftAgg+: Achieving Asymptotically Optimal Communication Load in Secure Aggregation for Federated Learning," *arXiv preprint arXiv:2203.13060*, 2022.

[14] K. Wan, H. Sun, M. Ji, and G. Caire, "Information Theoretic Secure Aggregation with Uncoded Groupwise Keys," *arXiv preprint arXiv:2204.11364*, 2022.

[15] Z. Wang and S. Ulukus, "Private Federated Submodel Learning via Private Set Union," *arXiv preprint arXiv:2301.07686*, 2023.

[16] Y. Zhao and H. Sun, "MDS Variable Generation and Secure Summation with User Selection," *arXiv preprint arXiv:2211.01220*, 2022.

[17] Y. J. Cho, J. Wang, and G. Joshi, "Client Selection in Federated Learning: Convergence Analysis and Power-of-Choice Selection Strategies," *arXiv preprint arXiv:2010.01243*, 2020.

[18] M. S. E. Mohamed, W.-T. Chang, and R. Tandon, "Privacy Amplification for Federated Learning via User Sampling and Wireless Aggregation," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3821–3835, 2021.

[19] S. Wang and M. Ji, "A Unified Analysis of Federated Learning with Arbitrary Client Participation," *arXiv preprint arXiv:2205.13648*, 2022.

Authorized licensed use limited to: University of North Texas. Downloaded on August 31,2023 at 19:54:48 UTC from IEEE Xplore. Restrictions apply.