Private Sum Computation: Trade-Off between Shared Randomness and Privacy

Rémi A. Chou
CSE Department
University of Texas at Arlington
Arlington, TX
remi.chou@uta.edu

Jörg Kliewer
ECE Department
New Jersey Institute of Technology
Newark, NJ
jkliewer@njit.edu

Aylin Yener
ECE Department
The Ohio State University
Columbus, OH
yener@ece.osu.edu

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619082

***Abstract*—Consider a scenario involving multiple users and a fusion center. Each user possesses a sequence of bits and can communicate with the fusion center through a one-way public channel. The fusion center's task is to compute the sum of all the sequences under the privacy requirement that a set of colluding users, along with the fusion center, cannot gain more than a predetermined amount δ of information, measured through mutual information, about the sequences of other users. Our first contribution is to characterize the minimum amount of necessary communication between the users and the fusion center, as well as the minimum amount of necessary shared randomness at the users. Our second contribution is to establish a connection between secure summation and secret sharing by showing that secret sharing is necessary to generate the local randomness needed for private summation, and prove that it holds true for any $\delta \geqslant 0$.**

I. Introduction

The study of distributed summation under security constraints is closely related to the problem of secure aggregation, as evidenced by prior works such as [1]–[10]. This area finds applications in distributed computing. Notably, secure summation with zero information leakage has been extensively explored in [5], [11]–[14].

In this paper, we extend the scope of secure sum computation to allow a controlled amount of information leakage. Specifically, we consider $L \geqslant 2$ users, each owning a sequence, and a fusion center tasked with computing the sum of these sequences. Our privacy requirement is to ensure that any colluding set of users of size T, where $T \leqslant L-2$ is fixed, along with the fusion center, must not learn more than a predetermined amount δ of information about the other users' sequences. Furthermore, inspired by ramp secret sharing, e.g., [15], [16], we also introduce a private summation setting where information leakage depends solely on the size of the colluding user set and increases linearly with its size.

Our first contribution is deriving converse results on the necessary communication rate of individual users to the fusion center and the required rate of shared randomness among users. These results are primarily established through combinatorial arguments. We also provide an achievability scheme that simultaneously matches all individual converse bounds. Our second contribution establishes a fundamental connection between secret sharing and private summation. When $\delta = 0$, it had been established, e.g., [17] that secret sharing can be employed to generate the local randomness at the users for secure summation. In this study, we establish a stronger connection between secure summation and secret sharing by showing that secret sharing is necessary to generate the local randomness needed for private summation, and prove that it also holds true for any $\delta \neq 0$.

Related work: The closest related work is [17], which corresponds to a special case where no information leakage is allowed, and communication rate and local randomness rate must be equal for all users. Other studies have explored secure computation with interactive communication from arbitrarily correlated randomness in various settings, e.g., [18]–[21]. We note that computation from correlated randomness that allows information leakage is also studied in [22], [23].

The remainder of the paper is organized as follows. The problem statement is presented in Section II. Our main results are summarized in Section III. The proofs of our converse results are presented in Sections IV and V. Finally, concluding remarks are presented in Section VI.

II. Problem Statement

Notation: Let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{Q}$ be the sets of natural, real, and rational numbers, respectively. For $a, b \in \mathbb{R}$, define $[\![a, b]\!] \triangleq [\lfloor a \rfloor, \lceil b \rceil] \cap \mathbb{N}$, and $[a] \triangleq [\![1, a]\!]$.

Consider a fusion center and L users who have individual sequences. The users communicate with the fusion center over a one-way, public, and noiseless channel, with the aim that the fusion center computes the sum of their sequences as described next.

Definition 1. *An $(L, n, (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ private-sum computation protocol consists of*

- *L users indexed in the set $[L]$;*
- *L independent sequences $(S_l)_{l\in[L]}$, where Sequence S_l is owned by User $l \in [L]$ and is uniformly distributed over the finite field $\mathbb{F}_2^n$;*
- *A source of global randomness shared by the L users, independent of the sequences $(S_l)_{l\in[L]}$, and described by U uniformly distributed over $\mathbb{F}_2^{nR^{(U)}}$;*

This work was supported in part by NSF grants CCF-2201824 and CCF-2201825.

- *L encoding functions $e_l : \mathbb{F}_2^{nR^{(U)}} \to \mathbb{F}_2^{nR_l^{(K)}}$;*
- *L encoding functions $e_l^{(X)} : \mathbb{F}_2^{nR_l^{(K)}} \times \mathbb{F}_2^n \to \mathbb{F}_2^{nR_l^{(X)}}$;*
- *A decoding function $d : \bigtimes_{l\in[L]} \mathbb{F}_2^{nR_l^{(X)}} \to \mathbb{F}_2^n$;*

and operates as follows:

1) *User $l \in [L]$ forms the local randomness $K_l \triangleq e_l(U)$;*
2) *User $l \in [L]$ encodes the sequence S_l as $X_l \triangleq e_l^{(X)}(K_l, S_l)$ and sends X_l over the public channel;*
3) *The fusion center computes $\hat{\Sigma}_{[L]} \triangleq d(X_{[L]})$ an estimate of $\Sigma_{[L]} \triangleq \sum_{l\in[L]} S_l$, where $X_{[L]} \triangleq (X_l)_{l\in[L]}$.*

Then, the desired requirements for an $(L, n, (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ private-sum computation protocol defined as in Definition 1 are as follows.

Definition 2. *Let $\delta \in \mathbb{Q}_+$ and $T \in \llbracket 0, L-2 \rrbracket$. An $(L, n, (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ private-sum computation protocol is (δ, T)-private if*

$$\hat{\Sigma}_{[L]} = \Sigma_{[L]}, \tag{1}$$

$$\max_{\substack{\mathcal{T}\subset[L]\\ :|\mathcal{T}|\leqslant T}} I(S_{[L]}; X_{[L]} | \Sigma_{[L]} S_{\mathcal{T}} K_{\mathcal{T}}) \leqslant \delta, \tag{2}$$

where for any $\mathcal{T} \subset [L]$, $S_{\mathcal{T}} \triangleq (S_l)_{l\in\mathcal{T}}$ and $K_{\mathcal{T}} \triangleq (K_l)_{l\in\mathcal{T}}$.

Equation (1) means that the fusion center computes the sum $\Sigma_{[L]}$ without errors. If the fusion center and a set $\mathcal{T}$ of T users collude, then, Equation (2) quantifies the amount of information that $X_{\mathcal{T}^c}$ leaks about $S_{\mathcal{T}^c}$ given the knowledge of $(\Sigma_{[L]}, S_{\mathcal{T}}, K_{\mathcal{T}})$.

In this study, for protocols that satisfy the requirements of Definition 2, we are interested in characterizing (i) the minimum amount of global randomness U needed, (ii) the minimum amount of public communication needed for each user, (iii) the minimum amount of local randomness needed for the users. To this end, we introduce the following quantities.

Definition 3. *Let $\delta \in \mathbb{Q}_+$ and $T \in [0, K-2] \cap \mathbb{N}$. Let $\mathcal{C}$ be the set of tuples $\Lambda \triangleq (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ such that there exist $(L, n, (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ private-sum computation protocols that are (δ, T)-private. Then, define*

$$R_{l,\star}^{(X)} \triangleq \inf_{\Lambda\in\mathcal{C}} R_l^{(X)}, \forall l \in [L],$$

$$R_{l,\star}^{(K)} \triangleq \inf_{\Lambda\in\mathcal{C}} R_l^{(K)}, \forall l \in [L],$$

$$R_{\Sigma,\star}^{(K)} \triangleq \inf_{\Lambda\in\mathcal{C}} \sum_{l\in[L]} R_l^{(K)},$$

$$R_\star^{(U)} \triangleq \inf_{\Lambda\in\mathcal{C}} R^{(U)}.$$

Remark. *In the special case $\delta = 0$, and when the local randomness rates are assumed identical for all users (i.e., $\exists C_1 \in \mathbb{R}, R_l^{(K)} = C_1, \forall l \in [L]$), and the communication rates are assumed identical for all users (i.e., $\exists C_2 \in \mathbb{R}, R_l^{(X)} = C_2, \forall l \in [L]$), then our model recovers the setting formalized in [17].*

Remark. *In Definition 2, δ is a rational number. However, note that by density of $\mathbb{Q}$ in $\mathbb{R}$, for any $\delta' \in \mathbb{R}_+$, for any $\epsilon > 0$, there exists $\delta \in \mathbb{Q}_+$ such that $|\delta - \delta'| \leqslant \epsilon$.*

Remark. *In Definition 2, it is sufficient to consider $T \in [0, L-2] \cap \mathbb{N}$. If $T = L$, then (2) is always satisfied as the left-hand side is equal to zero. This is also true if $T = L-1$, as $S_{[L]}$ can be reconstructed from $(\Sigma_{[L]}, S_{\mathcal{T}})$ for any $\mathcal{T} \subset [L]$ such that $|\mathcal{T}| = T$.*

III. Main results

We first show in Section III-A that Definition 2 can be simplified without loss of generality. We then present our converse and capacity results in Sections III-B and III-C, respectively. Finally, we establish the connection between secret sharing and shared randomness in private summation in Section III-D.

A. Preliminaries

One can prove that in the privacy constraint (2) of Definition 2, it is sufficient to consider δ of the form $\delta = n\alpha(L-1)$ with $\alpha \in [0,1] \cap \mathbb{Q}$, i.e., (2) can be replaced by

$$\max_{\substack{\mathcal{T}\subset[L]\\ :|\mathcal{T}|\leqslant T}} I(S_{[L]}; X_{[L]} | \Sigma_{[L]} S_{\mathcal{T}} K_{\mathcal{T}}) \leqslant n\alpha(L-1). \tag{3}$$

B. Converse results

Theorem 1 (Converse)**.** *For any $(L, n, (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ private-sum computation protocol that is $(\delta = n\alpha(L-1), T)$-private, $\alpha \in [0,1] \cap \mathbb{Q}$, we have*

$$R_l^{(X)} \geqslant 1, \forall l \in [L], \tag{4}$$

$$\textstyle\sum_{l\in[L]} R_l^{(K)} \geqslant (1-\alpha)L, \tag{5}$$

$$R^{(U)} \geqslant (1-\alpha)(L-1). \tag{6}$$

Proof. See Section IV. ∎

We now derive a converse on the individual rate of local randomness under the following leakage symmetry assumption:

$$\forall l \in \llbracket 0, L \rrbracket, \exists C_l \in \mathbb{R}_+, \forall \mathcal{T} \subseteq [L],$$
$$|\mathcal{T}| = l \implies I(S_{[L]}; X_{[L]} | \Sigma_{[L]} S_{\mathcal{T}} K_{\mathcal{T}}) = C_l, \tag{7a}$$

$$(C_l)_{l\in\llbracket 0, L-1\rrbracket} \text{ constantly decreases from } n\alpha(L-1) \text{ to } 0. \tag{7b}$$

Equation (7a) means that information leakage to a set of colluding users solely depends on the size of this set, rather than being influenced by the individual identities within the set. Equation (7b) means that the information leakage $I(S_{[L]}; X_{[L]} | \Sigma_{[L]} S_{\mathcal{T}} K_{\mathcal{T}})$ constantly decreases from its maximal value $n\alpha(L-1)$ to its minimal value 0, as the size of the set $\mathcal{T}$ increases from 0 to $L-1$. In other words, there is a uniform consideration of all users, rendering them indistinguishable in terms of their capacity to gain information about other users. We remark that the leakage symmetry condition (7) is similar to the leakage symmetry assumption defined for ramp secret sharing, e.g., [15], [16]. Indeed, (7) implies that for any $l \in [\![0, L-1]\!]$, $C_l = n\alpha(L-l-1)$.

Theorem 2 (Converse)**.** *Suppose that leakage symmetry, i.e., (7), is required. Then, for any $(L, n, (R_l^{(X)})_{l\in[L]}, R^{(U)}, (R_l^{(K)})_{l\in[L]})$ private-sum computation protocol that is $(\delta = n\alpha(L-1), T)$-private, $\alpha \in [0,1] \cap \mathbb{Q}$, we have*

$$\begin{aligned} R_l^{(X)} &\geqslant 1, \forall l \in [L], \\ R_l^{(K)} &\geqslant 1-\alpha, \forall l \in [L], \\ R^{(U)} &\geqslant (1-\alpha)(L-1). \end{aligned}$$

Proof. See Section V. ∎

C. Capacity results

Theorem 3. *For any $\alpha \in [0,1] \cap \mathbb{Q}$, we have*

$$\begin{aligned} R_{l,\star}^{(K)} &= (1-\alpha), \forall l \in [L], \\ R_{\Sigma,\star}^{(K)} &= L(1-\alpha), \text{ when (7) holds}, \\ R_{l,\star}^{(X)} &= 1, \forall l \in [L], \\ R_{\star}^{(U)} &= (1-\alpha)(L-1). \end{aligned}$$

Moreover, there exists a private-sum computation protocol that is $(\delta = n\alpha(L-1), T)$-private and simultaneously achieves $(R_{l,\star}^{(X)})_{l\in[L]}$, $R_{\star}^{(U)}$, $(R_{l,\star}^{(K)})_{l\in[L]})$, $R_{\Sigma,\star}^{(K)}$.

The achievability proof can be obtained as a time-sharing version of the coding scheme in [17], which proves that the converse bounds of Section III-B are tight. The details are omitted due to space constraints.

D. Connection between secret sharing and private summation

1) Preliminaries: We first review the notion of uniform secret sharing with leakage, e.g., [24]–[27].

Definition 4 (Uniform secret sharing with leakage)**.** *Let $\alpha \in [0,1] \cap \mathbb{Q}$, $t \in [L]$ and $z \in [t-1]$. An (α, t, z)- secret sharing scheme consists of*

- *A secret S uniformly distributed over $\{0,1\}^{n_s}$;*
- *A stochastic encoder $e : \{0,1\}^{n_s} \times \{0,1\}^{n_r} \to \{0,1\}^{n_{sh}}, (S, R) \mapsto (H_l)_{l\in[L]}$, which takes as input the secret S and a randomization sequence R uniformly distributed over $\{0,1\}^{n_r}$ and independent of S, and outputs L shares $(H_l)_{l\in[L]}$, with sum length n_{sh}, that are not necessarily of same length. For any $\mathcal{S} \subseteq [L]$, we define $H_{\mathcal{S}} = (H_l)_{l\in\mathcal{S}}$;*

and satisfies the two conditions

$$\max_{\mathcal{T}\subseteq[L]:|\mathcal{T}|=t} H(S|H_{\mathcal{T}}) = 0, \quad \textit{(Recoverability)} \tag{8}$$

$$\max_{\mathcal{U}\subseteq[L]:|\mathcal{U}|\leqslant z} I(S; H_{\mathcal{U}}) \leqslant \alpha H(S), \quad \textit{(Privacy leakage)} \tag{9}$$

and the leakage symmetry condition

$$\forall l \in [L], \exists C_l \in \mathbb{R}^+, \forall \mathcal{T} \in [L], |\mathcal{T}| = l \implies \frac{I(S; H_{\mathcal{T}})}{H(S)} = C_l. \tag{10}$$

As an example, a (t, Δ, L) ramp secret sharing scheme, e.g., [15], [16], is a coding scheme as in Definition 4 with $\alpha = 0$, $z = t - \Delta$, and $C_l \triangleq \begin{cases} 0 & \text{if } l \in [\![0, t-\Delta]\!] \\ \frac{l-t+\Delta}{\Delta} & \text{if } l \in [\![t-\Delta+1, t]\!], \\ 1 & \text{if } l \in [\![t+1, L]\!] \end{cases}$ meaning that any t shares can reconstruct S, any set of shares less than or equal to $t-\Delta$ does not leak any information about S, and for sets of shares with cardinality in $[\![t-\Delta+1, t]\!]$, the leakage increases linearly with the set cardinality.

2) Private summation and ramp secret sharing: While it was known that secret sharing can be employed to generate the local randomness at the users for secure summation, e.g., [17], Theorem 4 establishes a stronger connection between secure summation and secret sharing by showing that secret sharing is necessary to generate the local randomness needed for secure summation. The proof is omitted due to space constraints.

Theorem 4 (Converse)**.** *Suppose that leakage symmetry, i.e., (7) is required, and consider an optimal private-sum computation protocol, i.e., for all $l \in [L]$, $R_l^{(K)} = (1-\alpha)$, $R_l^{(X)} = 1$, and $R^{(U)} = (1-\alpha)(L-1)$. Then, $(K_l)_{l\in[L]}$ must be the shares of a $(L-1, L-1, L)$ ramp secret sharing scheme where U is the secret.*

IV. Proof of Theorem 1

A. Communication rate converse: Proof of (4)

For any $l \in [L]$, we have

$$\begin{aligned} &nR_l^{(X)} \\ &\overset{(a)}{\geqslant} H(X_l) \\ &\overset{(b)}{\geqslant} H(X_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &\geqslant I(X_l; \Sigma_{[L]}|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &= H(\Sigma_{[L]}|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - H(\Sigma_{[L]}|X_l S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &\overset{(c)}{=} H(\Sigma_{[L]}S_{[L]\setminus\{l\}}|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &\qquad - H(\Sigma_{[L]}|X_l S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &\overset{(d)}{\geqslant} H(S_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - H(\Sigma_{[L]}|X_l S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &\overset{(e)}{=} H(S_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - H(\Sigma_{[L]}|X_{[L]} S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) \\ &\overset{(f)}{=} H(S_l) \\ &\overset{(g)}{=} n, \end{aligned}$$

where (a) holds by Definition 1, (b) holds because conditioning reduces entropy, (c) holds by the chain rule, (d) holds by the chain rule and because S_l can be reconstructed from $(\Sigma_{[L]}, S_{[L]\setminus\{l\}})$, (e) holds because $X_{[L]\setminus\{l\}}$ is a function of $(S_{[L]\setminus\{l\}}, K_{[L]\setminus\{l\}})$, (f) holds by (1) and independence between S_l and $(S_{[L]\setminus\{l\}}, K_{[L]\setminus\{l\}})$, (g) holds by uniformity of S_l.

B. Sum-rate converse on local randomness: Proof of (5)

For any $i \in [L]$, we have

$$\alpha n(L-1)$$

$$\begin{aligned}
&\overset{(a)}{\geqslant} I(S_{[L]};X_{[L]}|\Sigma_{[L]})\\
&= \sum_{l\in[L]\setminus\{i\}} I(S_l;X_{[L]}|S_{[l-1]\setminus\{i\}}\Sigma_{[L]})\\
&\quad + I(S_i;X_{[L]}|S_{[L]\setminus\{i\}}\Sigma_{[L]})\\
&\overset{(b)}{=} \sum_{l\in[L]\setminus\{i\}} I(S_l;X_{[L]}|S_{[l-1]\setminus\{i\}}\Sigma_{[L]})\\
&\geqslant \sum_{l\in[L]\setminus\{i\}} I(S_l;X_l|S_{[l-1]\setminus\{i\}}\Sigma_{[L]})\\
&\overset{(c)}{=} \sum_{l\in[L]\setminus\{i\}} I(S_l;X_l|S_{[l-1]\setminus\{i\}}\Sigma_{[l:L]\cup\{i\}})\\
&= \sum_{l\in[L]\setminus\{i\}} [I(S_l;X_lS_{[l-1]\setminus\{i\}}|\Sigma_{[l:L]\cup\{i\}})\\
&\quad - I(S_l;S_{[l-1]\setminus\{i\}}|\Sigma_{[l:L]\cup\{i\}})]\\
&\overset{(d)}{=} \sum_{l\in[L]\setminus\{i\}} I(S_l;X_lS_{[l-1]\setminus\{i\}}|\Sigma_{[l:L]\cup\{i\}})\\
&= \sum_{l\in[L]\setminus\{i\}} [I(S_l;X_l|\Sigma_{[l:L]\cup\{i\}})\\
&\quad + I(S_l;S_{[l-1]\setminus\{i\}}|X_l\Sigma_{[l:L]\cup\{i\}})]\\
&\overset{(e)}{=} \sum_{l\in[L]\setminus\{i\}} I(S_l;X_l|\Sigma_{[l:L]\cup\{i\}})\\
&= \sum_{l\in[L]\setminus\{i\}} [I(S_l;X_l\Sigma_{[l:L]\cup\{i\}}) - I(S_l;\Sigma_{[l:L]\cup\{i\}})]\\
&\overset{(f)}{=} \sum_{l\in[L]\setminus\{i\}} I(S_l;X_l\Sigma_{[l:L]\cup\{i\}})\\
&\geqslant \sum_{l\in[L]\setminus\{i\}} I(S_l;X_l),
\end{aligned}\tag{11}$$

where (a) holds by (3), (b) holds because S_i can be recovered from $(S_{[L]\setminus\{i\}},\Sigma_{[L]})$, in (c) we used the notation $\Sigma_{[l:L]\cup\{i\}} \triangleq \sum_{j\in\{l,l+1,\dots,L\}\cup\{i\}} S_j$, (d) holds because $I(S_l;S_{[l-1]\setminus\{i\}}|\Sigma_{[l:L]\cup\{i\}}) \leqslant I(S_l\Sigma_{[l:L]\cup\{i\}};S_{[l-1]\setminus\{i\}}) = 0$ by independence of the users' sequences, (e) holds because $I(S_l;S_{[l-1]\setminus\{i\}}|X_l\Sigma_{[l:L]\cup\{i\}}) \leqslant I(S_l\Sigma_{[l:L]\cup\{i\}}X_l;S_{[l-1]\setminus\{i\}}) \leqslant I(S_l\Sigma_{[l:L]\cup\{i\}}K_l;S_{[l-1]\setminus\{i\}}) \leqslant I(S_l\Sigma_{[l:L]\cup\{i\}}U;S_{[l-1]\setminus\{i\}}) = 0$ by independence of the users' sequences and the global randomness, (f) holds by uniformity of the users' sequences and because $l \neq i$.

Hence, by remarking that

$$\sum_{i\in[L]}\sum_{l\in[L]\setminus\{i\}} I(S_l;X_l) = (L-1)\sum_{l\in[L]} I(S_l;X_l), \tag{12}$$

from (11) and (12), we have

$$\sum_{l\in[L]} I(S_l;X_l) \leqslant \alpha nL. \tag{13}$$

Then, we have

$$\begin{aligned}
\sum_{l\in[L]} nR_l^{(K)} &\overset{(a)}{\geqslant} \sum_{l\in[L]} H(K_l)\\
&\overset{(b)}{=} \sum_{l\in[L]} H(K_l|S_l)\\
&\geqslant \sum_{l\in[L]} I(K_l;X_l|S_l)\\
&= \sum_{l\in[L]} [H(X_l|S_l) - H(X_l|K_lS_l)]\\
&\overset{(c)}{=} \sum_{l\in[L]} H(X_l|S_l)\\
&= \sum_{l\in[L]} [H(X_l) - I(X_l;S_l)]\\
&\overset{(d)}{\geqslant} nL - \sum_{l\in[L]} I(X_l;S_l)\\
&\overset{(e)}{\geqslant} nL - \alpha nL\\
&= nL(1-\alpha),
\end{aligned}$$

where (a) holds by Definition 1, (b) holds by independence between U and $S_{[L]}$, (c) holds by Definition 1, (d) holds by (4), (e) holds by (13).

C. *Converse on global randomness: Proof of* (6)

We have

$$\begin{aligned}
&nR^{(U)}\\
&\geqslant H(U)\\
&\overset{(a)}{\geqslant} H(K_{[L]})\\
&\overset{(b)}{\geqslant} H(K_{[L]}|S_{[L]})\\
&\geqslant I(K_{[L]};X_{[L]}|S_{[L]})\\
&\overset{(c)}{=} H(X_{[L]}|S_{[L]})\\
&= H(X_{[L]}) - I(X_{[L]};S_{[L]})\\
&= H(X_{[L]}) - I(X_{[L]};\Sigma_{[L]}S_{[L]})\\
&= H(X_{[L]}) - I(X_{[L]};\Sigma_{[L]}) - I(X_{[L]};S_{[L]}|\Sigma_{[L]})\\
&\overset{(d)}{\geqslant} H(X_{[L]}) - I(X_{[L]};\Sigma_{[L]}) - \alpha n(L-1)\\
&= H(X_{[L]}) - H(\Sigma_{[L]}) + H(\Sigma_{[L]}|X_{[L]}) - \alpha n(L-1)\\
&\overset{(e)}{=} H(X_{[L]}) - n - \alpha n(L-1)\\
&= \sum_{l\in[L]} H(X_l|X_{[l-1]}) - n - \alpha n(L-1)\\
&\overset{(f)}{\geqslant} \sum_{l\in[L]} H(X_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}X_{[l-1]}) - n - \alpha n(L-1)\\
&\overset{(g)}{=} \sum_{l\in[L]} H(X_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - n - \alpha n(L-1)\\
&\geqslant \sum_{l\in[L]} I(X_l;\Sigma_{[L]}|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - n - \alpha n(L-1)\\
&= \sum_{l\in[L]} [H(\Sigma_{[L]}|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}})\\
&\quad - H(\Sigma_{[L]}|X_lS_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}})] - n - \alpha n(L-1)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(h)}{=} \sum_{l\in[L]} H(\Sigma_{[L]}|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - n - \alpha n(L-1)\\
&\overset{(i)}{=} \sum_{l\in[L]} H(\Sigma_{[L]}S_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - n - \alpha n(L-1)\\
&\geqslant \sum_{l\in[L]} H(S_l|S_{[L]\setminus\{l\}}K_{[L]\setminus\{l\}}) - n - \alpha n(L-1)\\
&\overset{(j)}{=} \sum_{l\in[L]} H(S_l) - n - \alpha n(L-1)\\
&\overset{(k)}{=} n(L-1)(1-\alpha),
\end{aligned}
$$

where (a) holds by Definition 1, (b) holds because conditioning reduces entropy, (c) holds because $X_{[L]}$ is a function of $(K_{[L]}, S_{[L]})$, (d) holds by (3), (e) holds by uniformity of the users' sequences and (1), (f) holds because conditioning reduces entropy, (g) holds because $X_{[l-1]}$ is a function of $(S_{[L]\setminus\{l\}}, K_{[L]\setminus\{l\}})$, (h) holds by (1) since $X_{[L]}$ can be reconstructed from $(X_l, S_{[L]\setminus\{l\}}, K_{[L]\setminus\{l\}})$, (i) holds by the chain rule, (j) holds by independence between the users' sequences and the local randomness, (k) holds by uniformity of the users' sequences.

V. Proof of Theorem 2

Lemma 1. *For any $\mathcal{T} \subseteq [L]$, we have*

$$C_{|\mathcal{T}|} = n(L - |\mathcal{T}| - \mathbb{1}\{\mathcal{T} \neq [L]\}) - H(S_{\mathcal{T}^c}|X_{[L]}K_{\mathcal{T}}S_{\mathcal{T}}).$$

The proof of Lemma 1 is omitted due to space constraints. Fix $l \in [L]$. For $t \in [\![0, L-1]\!]$, define

$$\mathcal{T}_t \triangleq \begin{cases} [\![1,t]\!] & \text{if } t < l \\ [\![1,t+1]\!]\setminus\{l\} & \text{if } t \geqslant l \end{cases},$$

and $\mathcal{T}_L \triangleq [L]$.

Lemma 2. *For $t \in [\![0, L-1]\!]$, we have*

$$
\begin{aligned}
&H(K_lS_l|K_{\mathcal{T}_t}S_{\mathcal{T}_t}X_{[L]})\\
&= n\mathbb{1}\{t \neq L-1\} + C_{t+1} - C_t + H(K_l|K_{\mathcal{T}_t}S_{[L]}X_{[L]}).
\end{aligned}
$$

Proof. For $t \in [\![0, L-2]\!]$, we have

$$
\begin{aligned}
&H(K_lS_l|K_{\mathcal{T}_t}S_{\mathcal{T}_t}X_{[L]})\\
&= H(K_lS_lS_{\mathcal{T}_t^c}|K_{\mathcal{T}_t}S_{\mathcal{T}_t}X_{[L]})\\
&\quad - H(S_{\mathcal{T}_t^c}|K_{\mathcal{T}_t\cup\{l\}}S_{\mathcal{T}_t\cup\{l\}}X_{[L]})\\
&= H(S_{\mathcal{T}_t^c}|K_{\mathcal{T}_t}S_{\mathcal{T}_t}X_{[L]}) + H(K_lS_l|K_{\mathcal{T}_t}S_{[L]}X_{[L]})\\
&\quad - H(S_{\mathcal{T}_t^c}|K_{\mathcal{T}_t\cup\{l\}}S_{\mathcal{T}_t\cup\{l\}}X_{[L]})\\
&\overset{(a)}{=} n(L-t-1) - C_t + H(K_l|K_{\mathcal{T}_t}S_{[L]}X_{[L]})\\
&\quad - H(S_{\mathcal{T}_t^c}|K_{\mathcal{T}_t\cup\{l\}}S_{\mathcal{T}_t\cup\{l\}}X_{[L]})\\
&= n(L-t-1) - C_t + H(K_l|K_{\mathcal{T}_t}S_{[L]}X_{[L]})\\
&\quad - H(S_{\mathcal{T}_t^c\setminus\{l\}}|K_{\mathcal{T}_t\cup\{l\}}S_{\mathcal{T}_t\cup\{l\}}X_{[L]})\\
&\overset{(b)}{=} n(L-t-1) - C_t + H(K_l|K_{\mathcal{T}_t}S_{[L]}X_{[L]}) + C_{t+1}\\
&\quad - n(L-(t+1)-1)\\
&= n + C_{t+1} - C_t + H(K_l|K_{\mathcal{T}_t}S_{[L]}X_{[L]}),
\end{aligned}
$$

where (a) and (b) hold by Lemma 1. Moreover, we have

$$
\begin{aligned}
&H(K_lS_l|K_{\mathcal{T}_{L-1}}S_{\mathcal{T}_{L-1}}X_{[L]})\\
&\overset{(a)}{=} H(K_l|K_{\mathcal{T}_{L-1}}S_{[L]}X_{[L]})\\
&\overset{(b)}{=} C_L - C_{L-1} + H(K_l|K_{\mathcal{T}_{L-1}}S_{[L]}X_{[L]}),
\end{aligned}
$$

where (a) holds by (1), (b) holds because $C_{L-1} = 0 = C_L$. ■

Then, we have

$$
\begin{aligned}
&H(K_l)\\
&\overset{(a)}{\geqslant} H(K_l) + n - H(X_l)\\
&\overset{(b)}{=} H(K_l) + H(S_l) - H(X_l)\\
&\overset{(c)}{=} H(K_lS_l) - H(X_l)\\
&\overset{(d)}{=} H(K_lS_lX_l) - H(X_l)\\
&= H(K_lS_l|X_l)\\
&\overset{(e)}{\geqslant} H(K_lS_l|K_{\mathcal{T}_0}S_{\mathcal{T}_0}X_{[L]})\\
&\geqslant H(K_lS_l|K_{\mathcal{T}_0}S_{\mathcal{T}_0}X_{[L]}) - H(K_lS_l|K_{\mathcal{T}_{L-1}}S_{\mathcal{T}_{L-1}}X_{[L]})\\
&= \sum_{i=0}^{L-2}[H(K_lS_l|K_{\mathcal{T}_i}S_{\mathcal{T}_i}X_{[L]}) - H(K_lS_l|K_{\mathcal{T}_{i+1}}S_{\mathcal{T}_{i+1}}X_{[L]})]\\
&\overset{(f)}{=} \sum_{i=0}^{L-3}[C_{i+1} - C_i - C_{i+2} + C_{i+1} + H(K_l|K_{\mathcal{T}_i}S_{[L]}X_{[L]})\\
&\quad - H(K_l|K_{\mathcal{T}_{i+1}}S_{[L]}X_{[L]})]\\
&\quad + [n + C_{L-1} - C_{L-2} + H(K_l|K_{\mathcal{T}_{L-2}}S_{[L]}X_{[L]}) - C_L\\
&\qquad + C_{L-1} - H(K_l|K_{\mathcal{T}_{L-1}}S_{[L]}X_{[L]})]\\
&\overset{(g)}{\geqslant} n + \sum_{i=0}^{L-2}[C_{i+1} - C_i - C_{i+2} + C_{i+1}]\\
&\overset{(h)}{=} n(1-\alpha),
\end{aligned}
$$

where (a) holds by (4), (b) holds by uniformity of the users' sequences, (c) holds by independence between the users' sequences and the local randomness, (d) holds by Definition 1, (e) holds because conditioning reduces entropy, (f) holds by Lemma 2, (g) holds because $H(K_l|K_{\mathcal{T}_i}S_{[L]}X_{[L]}) \geqslant H(K_l|K_{\mathcal{T}_{i+1}}S_{[L]}X_{[L]})$ for $i \in [\![0, L-2]\!]$ since conditioning reduces entropy, (h) follows from $C_{L-1} = 0 = C_L$ and (7).

VI. Concluding remarks

We have studied the problem of private summation from distributed users over a one-way, public, and noiseless channel. Our setting generalizes the problem of secure summation by allowing a controlled amount of information leakage for a given set of colluding users. We have characterized the minimum amount of required communication and shared randomness at the users. We have also established a strong connection between secret sharing and private summation by showing that secret sharing is necessary to generate the local randomness at the users.

References

[1] 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.

[2] J. H. Bell, K. A. Bonawitz, A. Gascón, T. Lepoint, and M. Raykova, “Secure single-server aggregation with (poly) logarithmic overhead,” in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 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 Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 479–489, 2021.

[4] 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.

[5] 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.

[6] J. So, C. He, C.-S. Yang, S. Li, Q. Yu, R. E Ali, B. Guler, and S. Avestimehr, “Lightsecagg: a lightweight and versatile design for secure aggregation in federated learning,” *Proceedings of Machine Learning and Systems*, vol. 4, pp. 694–720, 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 Journal on Selected Areas in Communications*, vol. 41, no. 4, pp. 977–989, 2023.

[8] K. Wan, H. Sun, M. Ji, and G. Caire, “Information theoretic secure aggregation with uncoded groupwise keys,” *arXiv preprint arXiv:2204.11364*, 2022.

[9] R. Schlegel, S. Kumar, E. Rosnes, and A. G. i Amat, “Codedpaddedfl and codedsecagg: Straggler mitigation and secure aggregation in federated learning,” *IEEE Transactions on Communications*, 2023.

[10] Z. Liu, J. Guo, K.-Y. Lam, and J. Zhao, “Efficient dropout-resilient aggregation for privacy-preserving machine learning,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1839–1854, 2022.

[11] B. Chor and E. Kushilevitz, “A communication-privacy tradeoff for modular addition,” *Information Processing Letters*, vol. 45, no. 4, pp. 205–210, 1993.

[12] M. Hayashi and T. Koshiba, “Secure modulo zero-sum randomness as cryptographic resource,” *Cryptology ePrint Archive*, 2018.

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

[14] K. Wan, H. Sun, M. Ji, and G. Caire, “On secure distributed linearly separable computation,” *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 3, pp. 912–926, 2022.

[15] G. R. Blakley and C. Meadows, “Security of ramp schemes,” in *Workshop on the Theory and Application of Cryptographic Techniques*. Springer, 1984, pp. 242–268.

[16] H. Yamamoto, “Secret sharing system using (k, l, n) threshold scheme,” *Electronics and Communications in Japan (Part I: Communications)*, vol. 69, no. 9, pp. 46–54, 1986.

[17] Y. Zhao and H. Sun, “Secure summation: Capacity region, groupwise key, and feasibility,” *arXiv preprint arXiv:2205.08458*, 2022.

[18] H. Tyagi, “Distributed function computation with confidentiality,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 691–701, 2013.

[19] D. Data, G. R. Kurri, J. Ravi, and V. M. Prabhakaran, “Interactive secure function computation,” *IEEE Transactions on Information Theory*, vol. 66, no. 9, pp. 5492–5521, 2020.

[20] D. Data, V. M. Prabhakaran, and M. M. Prabhakaran, “Communication and randomness lower bounds for secure computation,” *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 3901–3929, 2016.

[21] H. Tyagi, P. Narayan, and P. Gupta, “When is a function securely computable?” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6337–6350, 2011.

[22] W. Tu and L. Lai, “On function computation with privacy and secrecy constraints,” *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6716–6733, 2019.

[23] R. A. Chou and J. Kliewer, “Function computation without secure links: Information and leakage rates,” in *IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1223–1228.

[24] M. Yoshida, T. Fujiwara, and M. P. Fossorier, “Optimal uniform secret sharing,” *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 436–443, 2018.

[25] M. Yoshida, T. Fujiwara, and M. Fossorier, “Optimum general threshold secret sharing,” in *International Conference on Information Theoretic Security, ICITS*. Springer, 2012, pp. 187–204.

[26] R. A. Chou and J. Kliewer, “Secure distributed storage: Rate-privacy trade-off and XOR-based coding scheme,” in *IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 605–610.

[27] ——, “Secure distributed storage: Optimal trade-off between storage rate and privacy leakage,” in *IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 1324–1329.