2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton) | 979-8-3503-2814-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/Allerton58177.2023.10313393

Distributed Storage Over a Public Channel: Trade-off between Privacy and Shared Key Lengths

Maryam Keshvari
Department of EECS
Wichita State University
Wichita, KS 67260
mxkeshvari@shockers.wichita.edu

Rémi A. Chou
Department of CSE
The University of Texas at Arlington
Arlington, TX 76019
remi.chou@uta.edu

***Abstract*—Consider a user who wants to store a file across multiple servers such that at least t servers are needed to reconstruct the file and any z colluding servers cannot learn more than a fraction α of the file. Unlike traditional secret-sharing models that assume the availability of secure channels at no cost, we assume that the user can only transmit data to the servers through a public channel and that the user and each server share an individual secret key of length n bits. For a fixed key length n, we characterize the fundamental trade-off between the privacy leakage parameter α and the file length that the user can store in the servers. Furthermore, for the optimal trade-off between α and the file length, we determine (i) the minimum amount of local randomness needed by the user, (ii) the minimum amount of public communication from the user to the servers, and (iii) the minimum storage requirement at the servers.**

I. Introduction

Centralizing sensitive information storage puts the entire database at risk of being compromised, in the case of a data breach. Adopting a decentralized storage strategy can provide resilience against data breaches and avoid having a single point of entry for hackers. For instance, secret sharing [2], [3] offers an effective solution to securely store data among L servers, such that any coalition of $t \leq L$ servers can reconstruct the file by pooling their information but any coalition of at most $t-1$ compromised servers cannot learn any information, in an information-theoretic sense, about the file. Applications of secret sharing to secure distributed storage have been extensively studied in a variety of contexts [4]–[12]. As motivated in [13]–[16], the servers could be distinct cloud storage service providers, since companies may outsource data storage to reduce operating costs.

In this paper, we consider a user who wants to store a file across L servers such that (i) any coalition of $t \leq L$ servers who combine their information can reconstruct the file, and (ii) any z colluding servers cannot learn more than a fraction α of the file. This introduces a trade-off between α and the maximum file length, that the user can store. In our model, unlike traditional secret sharing models [2], [3], no information-theoretically secure channels are available cost-free. Instead, the user communicates with each server through a one-way public channel, and shares with each server a secret key, comprising a uniformly distributed sequence of n bits. Our main contribution is to establish the fundamental trade-off between the privacy leakage α and the file length that the user can store. Additionally, for the optimal trade-off between α and the file length, we determine the minimum amount of local randomness needed by the user, the minimum amount of public communication, and the minimum amount of storage requirements at the servers. As discussed in Section III-B, our study reveals two distinct regimes when $\alpha \geq \frac{z}{t}$ and when $\alpha < \frac{z}{t}$. Additionally, when $\alpha = 0$, i.e., in the absence of privacy leakage, our results recover those of the previous study in [1].

The most challenging aspect of our study is to establish converse results for the maximum file size that the user can store, the optimal amount of local randomness needed by the user, and the optimal amount of public communication between the user and servers. Unlike conventional secret sharing models, our model requires considering in our analysis shared secret keys, open communication between all parties, and a joint design of the share creation phase and the share distribution phase. Indeed, conventional secret sharing models treat these two phases separately, focusing only on the creation phase of the shares and assuming that the secure communication phase occurs through information-theoretically secure channels without any cost. Then, we propose an achievable scheme that separates the distribution of the shares to the servers (via a one-time pad) from the generation of the shares (using the secret sharing scheme from [23]), and establish the optimality of this coding strategy by showing that its performance matches our converse results.

The paper is organized as follows. Section II provides a formal statement of the problem. Section III presents our main results. Finally, Section IV provides concluding remarks. Some proofs are omitted due to space constraints.

II. Problem Statement

Notation: For any $a \in \mathbb{N}^*$, define $[\![1, a]\!] \triangleq [1, a] \cap \mathbb{N}$. For $x \in \mathbb{R}$, define $[x]^+ \triangleq \max(0, x)$. For a given set $\mathcal{S}$, let $2^{\mathcal{S}}$ denote the power set of $\mathcal{S}$. Let also $\bigtimes$ denote the Cartesian product.

Consider L servers indexed by $\mathcal{L} \triangleq [\![1, L]\!]$ and one user. Assume that Server $l \in \mathcal{L}$ and the user share a secret key $K_l \in \mathcal{K} \triangleq \{0,1\}^n$, which is a sequence of n bits uniformly

This work was supported in part by NSF grant CCF-2201825.

distributed over $\{0,1\}^n$. The L keys are assumed to be jointly independent. For any $\mathcal{Y} \subseteq \mathcal{L}$, we use the notation $K_{\mathcal{Y}} \triangleq (K_y)_{y\in\mathcal{Y}}$.

Definition 1. *A $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private storage strategy consists of*

- A file F owned by the user, which is uniformly distributed over $\mathcal{F} \triangleq \{0,1\}^{r^{(F)}}$ and independent from the keys $K_{\mathcal{L}}$;
- A sequence of local randomness R owned by the user, which is uniformly distributed over $\mathcal{R} \triangleq \{0,1\}^{r^{(R)}}$ and independent from all other random variables;
- L encoding functions $h_l : \mathcal{R} \times \mathcal{K} \times \mathcal{F} \to \mathcal{M}_l$, where $l \in \mathcal{L}$ and $\mathcal{M}_l \triangleq \{0,1\}^{r_l^{(M)}}$;
- L servers with storage space $r_l^{(S)}$ bits for Server $l \in \mathcal{L}$;
- L encoding functions $g_l : \mathcal{M}_l \times \mathcal{K} \to \mathcal{S}_l$, where $l \in \mathcal{L}$ and $\mathcal{S}_l \triangleq \{0,1\}^{r_l^{(S)}}$;
- 2^L decoding function $f_{\mathcal{A}} : \bigtimes_{l\in\mathcal{A}} \mathcal{S}_l \to \mathcal{F}$, where $\mathcal{A} \subseteq \mathcal{L}$;

and operates as follows:

1) The user publicly sends to Server $l \in \mathcal{L}$ the message $M_l \triangleq h_l(R, K_l, F)$. We use the notation $M \triangleq (M_l)_{l\in\mathcal{L}}$.
2) Server $l \in \mathcal{L}$ stores $S_l \triangleq g_l(M_l, K_l)$.
3) Any subset of servers $\mathcal{A} \subseteq \mathcal{L}$ can compute $\hat{F}(\mathcal{A}) \triangleq f_{\mathcal{A}}(S_{\mathcal{A}})$ an estimate of F, where $S_{\mathcal{A}} \triangleq (S_l)_{l\in\mathcal{A}}$.

Definition 2. *Fix $t \in [\![1, L]\!]$, $z \in [\![1, t-1]\!]$ and $\alpha \in [0,1] \cap \mathbb{Q}$. Then, $r^{(F)}$ is (α, t, z)-achievable if there exists a $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private file storage strategy such that*

$$\max_{\forall \mathcal{A}\subseteq\mathcal{L}:|\mathcal{A}|\geq t} H(F|\hat{F}(\mathcal{A})) = 0, \quad \textit{(Reliability)} \tag{1}$$

$$\max_{\forall \mathcal{U}\subseteq\mathcal{L}:|\mathcal{U}|\leq z} \frac{I(F; M, K_{\mathcal{U}})}{H(F)} \leq \alpha, \quad \textit{(Privacy leakage)} \tag{2}$$

$$I(F; M) = 0. \quad \textit{(Security)} \tag{3}$$

The set of all (α, t, z)-achievable lengths $r^{(F)}$ is denoted by $\mathcal{C}_F(\alpha, t, z)$.

(1) means that any subset of servers with size greater than or equal to t is able to perfectly recover the file F, (2) means that any subset of servers with size smaller than or equal to z can learn at most a fraction α of the file, and (3) means that the public communication does not leak any information about the file. For a fixed α, our main objective is to determine, under the constraints (1)-(3), the maximal file length that the user can store in the servers given that the secret keys shared with the servers have length n. Next, another of our objectives is to determine (i) the minimum amount of local randomness at the user, (ii) the minimum storage requirement at the servers, and (iii) the minimum amount of public communication from the user to the servers that are needed to achieve the largest file rate in $\mathcal{C}_F(\alpha, t, z)$. To this end, we introduce the following definition.

Definition 3. *Fix $\alpha \in [0,1] \cap \mathbb{Q}$, $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. For $r^{(F)}$ in $\mathcal{C}_F(\alpha, t, z)$, let $\mathcal{Q}(r^{(F)})$ be the set of tuples $T \triangleq (r^{(R)}, (r_l^{(M)})_{l\in\mathcal{L}}, (r_l^{(S)})_{l\in\mathcal{L}})$ such that there exists a $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private file storage strategy that (α, t, z)-achieves $r^{(F)}$. Then, define*

$$r_*^{(F)}(\alpha, t, z) \triangleq \sup_{r^{(F)}\in\mathcal{C}_F(\alpha,t,z)} r^{(F)}, \tag{4}$$

$$r_{l,*}^{(M)}(\alpha, t, z) \triangleq \inf_{T\in\mathcal{Q}(r_*^{(F)}(\alpha,t,z))} r_l^{(M)}, l \in \mathcal{L}, \tag{5}$$

$$r_{\sum,*}^{(M)}(\alpha, t, z) \triangleq \inf_{T\in\mathcal{Q}(r_*^{(F)}(\alpha,t,z))} \sum_{l\in\mathcal{L}} r_l^{(M)}, \tag{6}$$

$$r_*^{(R)}(\alpha, t, z) \triangleq \inf_{T\in\mathcal{Q}(r_*^{(F)}(\alpha,t,z))} r^{(R)}, \tag{7}$$

$$r_{l,*}^{(S)}(\alpha, t, z) \triangleq \inf_{T\in\mathcal{Q}(r_*^{(F)}(\alpha,t,z))} r_l^{(S)}, l \in \mathcal{L}. \tag{8}$$

$r_*^{(F)}(\alpha, t, z)$ is the largest file size that the user can privately store under the constraints (1)-(3). Then, $r_*^{(R)}(\alpha, t, z)$, $r_{l,*}^{(M)}(\alpha, t, z)$, $r_{\sum,*}^{(M)}(\alpha, t, z)$, and $r_*^{(S)}(\alpha, t, z)$, $l \in \mathcal{L}$, correspond to the least amount of local randomness, the minimum amount of public communication to Server l, the minimum amount of public communication to all the servers, and the minimum storage size required at Server l, respectively, needed for the user to achieve $r_*^{(F)}(\alpha, t, z)$.

III. Main Results

A. Impossibility results

Let $\alpha \in [0,1] \cap \mathbb{Q}$, $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$.

Theorem 1. *(Converse on the file length). We have*

$$r_*^{(F)}(\alpha, t, z) \leq \min\left(\frac{t-z}{1-\alpha}, t\right) n.$$

Proof. See Appendix A. □

Theorem 1 means that it is impossible for the user to store a file of length larger than $\min\left(\frac{t-z}{1-\alpha}, t\right) n$ bits.

Theorem 2. *(Converse on storage size requirement at the server). We have*

$$r_{l,*}^{(S)}(\alpha, t, z) \geq n, \forall l \in \mathcal{L}.$$

Proof. See Appendix B. □

Theorem 2 means that Server $l \in \mathcal{L}$ needs a storage capacity of at least n bits.

Theorem 3. *(Converse on the total amount of public communication to the servers). We have*

$$\max\left(\frac{1-\alpha}{t-z}, \frac{1}{t}\right) L \times r_*^{(F)}(\alpha, t, z) \leq r_{\sum,*}^{(M)}(\alpha, t, z). \tag{9}$$

Proof. See Appendix C. □

Theorem 3 means that it is impossible for the user to store a file of length $r_*^{(F)}(\alpha, t, z)$ if the public communication sum length to the servers is smaller than $\max\left(\frac{1-\alpha}{t-z}, \frac{1}{t}\right) L \times r_*^{(F)}(\alpha, t, z)$ bits.

Theorem 4. *(Converse on the amount of public communication to an individual server). Consider the following condition*

$$\forall \mathcal{U}, \mathcal{V} \subseteq \mathcal{L}, |\mathcal{U}| = |\mathcal{V}| \Rightarrow I(F; M_{\mathcal{U}}, K_{\mathcal{U}}) = I(F; M_{\mathcal{V}}, K_{\mathcal{V}}). \tag{10}$$

(10) *indicates that any two sets of colluding servers that have the same size have the same amount of information about the file F. If* (10) *holds, then we have*

$$\max\left(\frac{1-\alpha}{t-z}, \frac{1}{t}\right) \times r_*^{(F)}(\alpha, t, z) \leq r_{l,*}^{(M)}(\alpha, t, z), \forall l \in \mathcal{L}.$$

Proof. See Appendix D. □

Note that (9) corresponds to leakage symmetry and had already been introduced in the context of secret sharing under the denomination uniform secret sharing [22]. Under the condition (9), Theorem 4 means that it is impossible for the user to store a file of length $r_*^{(F)}$ if the public communication length to Server $l \in \mathcal{L}$ is smaller than $\max\left(\frac{1-\alpha}{t-z}, \frac{1}{t}\right) \times r_*^{(F)}(\alpha, t, z)$ bits.

Theorem 5. *(Converse on the amount of required local randomness at the users). We have*

$$r_*^{(R)}(\alpha, t, z) \geq \frac{[z-\alpha t]^+}{t-z} r_*^{(F)}(\alpha, t, z). \tag{11}$$

The proof of Theorem 5 is omitted due to space constraints. Theorem 5 means that it is impossible for the user to store a file of length $r_*^{(F)}(\alpha, t, z)$ if the amount of its local randomness is smaller than $\frac{[z-\alpha t]^+}{t-z} r_*^{(F)}(\alpha, t, z)$ bits.

B. Capacity result

Theorem 6. *Let $\alpha \in [0,1] \cap \mathbb{Q}$, $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. There exists a $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l \in \mathcal{L}}, (2^{r_l^{(S)}})_{l \in \mathcal{L}})$ private file storage strategy that (α, t, z)-achieves $r^{(F)}$such that,*

$$\begin{aligned} r^{(F)} &= \min\left(\frac{t-z}{1-\alpha}, t\right) n, \\ r^{(R)} &= \frac{[z-\alpha t]^+}{1-\alpha} n, \\ r_l^{(S)} &= n, \forall l \in \mathcal{L}, \\ r_l^{(M)} &= n, \forall l \in \mathcal{L}. \end{aligned}$$

Note that Theorem 6 recovers the result in [1] when $\alpha = 0$, i.e., in the absence of privacy leakage.

From Theorem 6, we observe two distinct regimes for α. When $\alpha \geq \frac{z}{t}$ (respectively $\alpha < \frac{z}{t}$), the user can store a file of size at most nt bits (respectively $\frac{t-z}{1-\alpha} n$ bits) such that any set of servers larger than or equal to t can reconstruct the file, and any set of servers smaller than or equal to z can learn at most a fraction α of the file. Moreover, if the user stores a file of length $\frac{t-z}{1-\alpha} n$ bits (respectively nt bits), then the optimal storage capacity at each server is n bits, the minimum amount of local randomness needed by the user is $\frac{z-\alpha t}{1-\alpha} n$ bits (respectively 0 bits), and the optimal amount of public communication from the user to all the servers is $L \times n$ bits.

C. Coding strategy for the achievability of Theorem 6

1) Review of secret sharing with privacy leakage: Let $\alpha \in [0,1] \cap \mathbb{Q}$, $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$.

Definition 4. *An (α, t, z)- secret sharing scheme consists of*

- A secret S uniformly distributed over $\{0,1\}^{n_s}$;
- A stochastic encoder $e : \{0,1\}^{n_s} \times \{0,1\}^{n_r} \to \{0,1\}^{n_{sh} L}, (S, R) \mapsto (H_l)_{l \in \mathcal{L}}$, 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 $(H_l)_{l \in \mathcal{L}}$ of length n_{sh}. For any $\mathcal{S} \subseteq \mathcal{L}$, we define $H_{\mathcal{S}} = (H_l)_{l \in \mathcal{S}}$;

and satisfies the two conditions

$$\max_{\mathcal{T} \subseteq \mathcal{L}: |\mathcal{T}| = t} H(S | H_{\mathcal{T}}) = 0, \quad \text{(Recoverability)} \tag{12}$$

$$\max_{\mathcal{U} \subseteq \mathcal{L}: |\mathcal{U}| \leq z} I(S; H_{\mathcal{U}}) \leq \alpha H(F). \quad \text{(Privacy leakage)} \tag{13}$$

Theorem 7. *([23]) For a fixed secret length n_s, there exists an (α, t, z)-secret sharing scheme such that the length of a share n_{sh} and the length of the randomization sequence n_r satisfy*

$$n_{sh} = \max\left(\frac{1-\alpha}{t-z}, \frac{1}{t}\right) n_s, \quad n_r = \frac{[z-\alpha t]^+}{t-z} n_s.$$

2) Achievability scheme for Theorem 6: Consider a file F such that $r^{(F)} = |F| = \min\left(\frac{t-z}{1-\alpha}, t\right) n$. Then, the user forms $(H_l)_{l \in \mathcal{L}}$ with an (α, t, z) secret sharing scheme taken from Theorem 7 applied to F. By Theorem 7, for $l \in \mathcal{L}$, the length of a share is $|H_l| = \max\left(\frac{1-\alpha}{t-z}, \frac{1}{t}\right) \times \min\left(\frac{t-z}{1-\alpha}, t\right) n = n$, and the length of the randomization sequence is $n_r = \frac{[z-\alpha t]^+}{t-z} \times \min\left(\frac{t-z}{1-\alpha}, t\right) n$. Hence, since $|K_l| = n, l \in \mathcal{L}$, the user can form $M_l \triangleq H_l \oplus K_l$ and publicly send it to Server l, where $\oplus$ denotes bitwise modulo-two addition. Upon receiving M_l, Server l stores $S_l \triangleq K_l \oplus M_l = H_l$. The analysis of the coding scheme is omitted due to space constraints.

IV. Concluding remarks

We considered storing a file across L servers, ensuring that any set of at least t servers can reconstruct the file and any set of z colluding servers cannot learn more than a fraction α of the file. Hence, our model introduces a trade-off between the privacy leakage parameter α and the file length that the user can store on the servers. In contrast to traditional secret sharing, our model did not assume the existence of information-theoretically secure channels. Instead, the user communicates with the servers over a public channel and shares a secret key of length n with each server. For a given n, we characterized the optimal trade-off between the file length that the user can store and the privacy leakage parameter α. Additionally, when this optimal trade-off is achieved, we established the minimum amount of local randomness needed by the user, the minimum amount of publication communication, and the minimum storage requirement at the servers.

Appendix A
Proof of Theorem 1

Consider an arbitrary $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private file storage strategy that (α,t,z)-achieves $r^{(F)}$.

Lemma 1. *We have* $r_*^{(F)}(\alpha,t,z) \leq \frac{n(t-z)}{1-\alpha}$.

Proof. Let $\mathcal{A}$,$\mathcal{U} \subseteq \mathcal{L}$ such that $|\mathcal{A}|= t$, $|\mathcal{U}|= z$, and $\mathcal{U} \subset \mathcal{A}$. We have

$$\begin{aligned} r^{(F)} &\overset{(a)}{=} H(F)\\ &= H(F|M,K_{\mathcal{U}}) + I(F;M,K_{\mathcal{U}})\\ &\overset{(b)}{\leq} H(F|M,K_{\mathcal{U}}) + \alpha H(F)\\ &= I(\hat{F}(\mathcal{A});F|M,K_{\mathcal{U}}) + H(F|M,K_{\mathcal{U}},\hat{F}(\mathcal{A})) + \alpha H(F)\\ &\overset{(c)}{=} I(\hat{F}(\mathcal{A});F|M,K_{\mathcal{U}})+\alpha H(F)\\ &\overset{(d)}{\leq} I(M,K_{\mathcal{A}};F|M,K_{\mathcal{U}})+\alpha H(F)\\ &= I(K_{\mathcal{A}};F|M,K_{\mathcal{U}})+\alpha H(F)\\ &\overset{(e)}{\leq} I(K_{\mathcal{A}};K_{\mathcal{L}},F,R|K_{\mathcal{U}})+\alpha H(F)\\ &\overset{(f)}{=} I(K_{\mathcal{A}};K_{\mathcal{L}},F|K_{\mathcal{U}})+\alpha H(F)\\ &\overset{(g)}{=} I(K_{\mathcal{A}};K_{\mathcal{L}}|K_{\mathcal{U}})+\alpha H(F)\\ &= H(K_{\mathcal{A}}|K_{\mathcal{U}})+\alpha H(F)\\ &\overset{(h)}{=} H(K_{\mathcal{A}\backslash\mathcal{U}})+\alpha H(F)\\ &\overset{(i)}{=} n(t-z)+\alpha H(F)\\ &= n(t-z)+\alpha r^{(F)}, \end{aligned}\tag{14}$$

where (a) holds by uniformity of the file F, (b) holds by (2), (c) holds by (1), (d) holds by the data processing inequality, (e) holds by the chain rule and the data processing inequality because M is a function of $(F,R,K_{\mathcal{L}})$, (f) holds by independence between R and $(F,K_{\mathcal{L}})$, (g) holds by independence between F and $K_{\mathcal{L}}$, (h) holds because $\mathcal{U} \subset \mathcal{A}$, (i) holds because the keys are uniformly distributed and $|\mathcal{A}\backslash\mathcal{U}|= t-z$.

Then, from (14) we have $r^{(F)} \leq \frac{n(t-z)}{1-\alpha}$. Finally, note that (14) is valid for any private file storage strategy and, in particular, for a file storage strategy that achieves $r_*^F(\alpha,t,z)$. □

By choosing, $\mathcal{U} \triangleq \emptyset$ in the proof of Lemma 1, one can also prove that $r_*^{(F)}(\alpha,t,z) \leq nt$.

Appendix B
Proof of Theorem 2

Consider an arbitrary $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private file storage strategy that (α,t,z)-achieves $r^{(F)}$.

Server $l \in \mathcal{L}$ must store the key K_l at the beginning of the protocol. Hence, we must have $r_{l,*}^{(S)}(\alpha,t,z) \geq |K_l|= n$.

Appendix C
Proof of Theorem 3

Consider an arbitrary $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private file storage strategy that (α,t,z)-achieves $r^{(F)}$.

Using Definition 1, (1), and (2), one can prove the following lemma.

Lemma 2. *For $\mathcal{T} \subseteq \mathcal{L}$ and $\mathcal{S} \subseteq \mathcal{L}\backslash\mathcal{T}$ such that $|\mathcal{T}|= z$ and $|\mathcal{S}|= t-z$, we have*

$$\sum_{l\in S} H(M_l) \geq (1-\alpha)H(F). \tag{15}$$

Then, we have

$$\begin{aligned} &\frac{L}{t-z}r^{(F)}\\ &\overset{(a)}{=} \frac{L}{t-z}H(F)\\ &\overset{(b)}{=} \gamma \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=z}} \sum_{\substack{\mathcal{S}\subseteq\mathcal{T}^c\\|\mathcal{S}|=t-z}} H(F)\\ &\overset{(c)}{\leq} \gamma\frac{1}{1-\alpha} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=z}} \sum_{\substack{\mathcal{S}\subseteq\mathcal{T}^c\\|\mathcal{S}|=t-z}} \sum_{l\in\mathcal{S}} H(M_l)\\ &\overset{(d)}{=} \gamma\frac{1}{1-\alpha} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=z}} \binom{L-z-1}{t-z-1} \sum_{l\in\mathcal{T}^c} H(M_l)\\ &\overset{(e)}{=} \gamma\frac{1}{1-\alpha}\binom{L-z-1}{t-z-1} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=L-z}} \sum_{l\in\mathcal{T}} H(M_l)\\ &\overset{(f)}{=} \gamma\frac{1}{1-\alpha}\binom{L-z-1}{t-z-1}\binom{L-1}{L-z-1} \sum_{l\in\mathcal{L}} H(M_l)\\ &= \frac{1}{1-\alpha}\sum_{l\in\mathcal{L}} H(M_l)\\ &\leq \frac{1}{1-\alpha}\sum_{l\in\mathcal{L}} r_l^{(M)}, \end{aligned}\tag{16}$$

where (a) holds by uniformity of F, (b) holds with $\gamma \triangleq \frac{L}{t-z}\binom{L}{z}^{-1}\binom{L-z}{t-z}^{-1}$, (c) holds by (15), (d) holds by [21, Lemma 3.2], (e) holds by a change of variable in the sum, (f) holds by a change of variable in the sum.

Since (16) is valid for any private file storage strategy, we have

$$L\frac{1-\alpha}{t-z}r_*^{(F)}(\alpha,t,z) \leq r_{\sum,*}^{(M)}(\alpha,t,z). \tag{17}$$

Similar to (17), one can prove using Lemma 2 with $\mathcal{T} \triangleq \emptyset$,

$$\frac{L}{t}r_*^{(F)}(\alpha,t,z) \leq r_{\sum,*}^{(M)}(\alpha,t,z).$$

Appendix D
Proof of Theorem 4

Consider an arbitrary $(2^{r^{(F)}}, 2^{r^{(R)}}, (2^{r_l^{(M)}})_{l\in\mathcal{L}}, (2^{r_l^{(S)}})_{l\in\mathcal{L}})$ private file storage strategy that (α,t,z)-achieves $r^{(F)}$. Fix $l \in \mathcal{L}$.

Using (10), (1), and (2), one can prove the following lemma.

Lemma 3. *Let $\mathcal{F}$ be the set of all the functions f:$[\![1, t-z+2]\!] \rightarrow [0,1]$ that are non-decreasing and such that $f(1) = \alpha H(F)$, $f(t-z+2) = f(t-z+1) = H(F)$. Then, we have*

$$H(M_l) \geq \min_{f \in \mathcal{F}} \sum_{i=1}^{t-z} [2f(i+1) - f(i) - f(i+2)]^+.$$

Then, for $f \in \mathcal{F}$ and f^+ the concave envelope of f, we have

$$\begin{aligned}
&\sum_{i=1}^{t-z} [2f(i+1) - f(i) - f(i+2)]^+ \\
&\overset{(a)}{\geq} \sum_{i=1}^{t-z} [2f^+(i+1) - f^+(i) - f^+(i+2)] \\
&= \sum_{i=1}^{t-z} [(f^+(i+1) - f^+(i)) - (f^+(i+2) - f^+(i+1))] \\
&= (f^+(2) - f^+(1)) - (f^+(t-z+2) - f^+(t-z+1)) \\
&\overset{(b)}{\geq} (f^+(t-z+1) - f^+(1))/(t-z) \\
&\overset{(c)}{=} (H(F) - \alpha H(F))/(t-z) \\
&= H(F)(1-\alpha)/(t-z), \qquad (18)
\end{aligned}$$

where (a) holds by [1, (17) and (18)], (b) holds because $f^+(t-z+2) = f^+(t-z+1) = H(F)$, and $f^+(2) - f^+(1) \geq (f^+(t-z+1) - f^+(1))/(t-z)$ by concavity of f^+, (c) holds because $f^+(t-z+1) = H(F)$ and $f^+(1) = \alpha H(F)$.

Then, by Lemma 3, the uniformity of F, and (18), we have

$$\begin{aligned}
r_l^{(M)} &\geq H(M_l) \\
&\geq r^{(F)} \frac{1-\alpha}{t-z}. \qquad (19)
\end{aligned}$$

Since (19) is valid for any private file storage strategy, we have

$$\frac{1-\alpha}{t-z} r_*^{(F)}(\alpha, t, z) \leq r_{l,*}^{(M)}(\alpha, t, z).$$

Similarly, by modifying Lemma 3, one can prove that

$$\frac{1}{t} r_*^{(F)}(\alpha, t, z) \leq r_{l,*}^{(M)}(\alpha, t, z).$$

References

[1] R. Chou, "Quantifying the cost of privately storing data in distributed storage systems," in IEEE Transactions on Information Theory, 2022. 970–975.

[2] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11, pp. 612–613, Nov. 1979.

[3] G. Blakley, "Safeguarding cryptographic keys," Proceedings of the National Computer Conference, pp. 313–317, 1979.

[4] A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, "Centralized repair of multiple node failures with applications to communication efficient secret sharing," IEEE Transactions on Information Theory, vol. 64, no. 12, pp. 7529–7550, 2018.

[5] A. Agarwal and A. Mazumdar, "Security in locally repairable storage," IEEE Transactions on Information Theory, vol. 62, no. 11, pp. 6204– 6217, 2016.

[6] M. Soleymani and H. Mahdavifar, "Distributed multi-the user secret sharing," IEEE Transactions on Information Theory, vol. 67, no. 1, pp. 164– 178, 2020.

[7] W. Huang, M. Langberg, J. Kliewer, and J. Bruck, "Communication efficient secret sharing," IEEE Transactions on Information Theory, vol. 62, no. 12, pp. 7195–7206, 2016.

[8] R. Bitar and S. El Rouayheb, "Staircase codes for secret sharing with optimal communication and read overheads," IEEE Transactions on Information Theory, vol. 64, no. 2, pp. 933–943, 2017.

[9] N. B. Shah, K. Rashmi, and K. Ramchandran, "Distributed secret dissemination across a network," IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 7, pp. 1206–1216, 2015.

[10] W. Huang and J. Bruck, "Secure RAID schemes for distributed storage," in IEEE International Symposium on Information Theory (ISIT). 2016, pp. 1401–1405.

[11] ——, "Secret sharing with optimal decoding and repair bandwidth," in IEEE International Symposium on Information Theory (ISIT). 2017, pp. 1813–1817.

[12] 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.

[13] A. Bessani, M. Correia, B. Quaresma, F. Andre, and P. Sousa, "DepSky: ´ Dependable and secure storage in a cloud-of-clouds," ACM Transactions on Storage, vol. 9, no. 4, pp. 1–33, 2013.

[14] R. Shor, G. Yadgar, W. Huang, E. Yaakobi, and J. Bruck, "How to best share a big secret," in Proceedings of the 11th ACM International Systems and Storage Conference, 2018, pp. 76–88.

[15] B. Fabian, T. Ermakova, and P. Junghanns, "Collaborative and secure sharing of healthcare data in multi-clouds," Information Systems, vol. 48, pp. 132–150, 2015.

[16] W. Huang and J. Bruck, "Secure RAID schemes from EVEN-ODD and STAR codes," in IEEE International Symposium on Information Theory (ISIT). 2017, pp. 609–613.

[17] E. Karnin, J. Greene, and M. Hellman, "On secret sharing systems," IEEE Transactions on Information Theory, vol. 29, no. 1, pp. 35–41, 1983.

[18] R. J. McEliece and D. V. Sarwate, "On sharing secrets and ReedSolomon codes," Communications of the ACM, vol. 24, no. 9, pp. 583– 584, 1981.

[19] J. Benaloh and J. Leichter, "Generalized secret sharing and monotone functions," in Conference on the ory and Application of Cryptography. Springer, 1988, pp. 27–35.

[20] M. Ito, A. Saito, and T. Nishizeki, "Secret sharing scheme realizing general access structure," Electronics and Communications in Japan (Part III: Fundamental Electronic Science), vol. 72, no. 9, pp. 56–64, 1989.

[21] C. Blundo, A. De Santis, and U. Vaccaro, "Randomness in distribution protocols," Information and Computation, vol. 131, no. 2, pp. 111–139, 1996.

[22] G. R. Blakley and C. Meadows, "Security of ramp schemes," in Workshop on the ory and Application of Cryptographic Techniques. Springer, 1984, pp. 242–268.

[23] R. Chou, J. Kliewer, "Secure Distributed Storage: Optimal Trade-Off Between Storage Rate and Privacy Leakage" in Proc. of the IEEE International Symposium on Information Theory (ISIT). 2023