Quantifying the Cost of Privately Storing Data in Distributed Storage Systems

Rémi A. Chou
Department of Electrical Engineering & Computer Science
Wichita State University
Wichita, KS 67260
remi.chou@wichita.edu

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834900

***Abstract*—Consider a user who wishes to store a file in multiple servers such that at least t servers are needed to reconstruct the files, and z colluding servers cannot learn any information about the file. Unlike traditional models, where perfectly secure channels are assumed to be available at no cost between the user and each server, we assume that the user can only send data to the servers via public channels, and that the user and each server share an individual secret key with length n. For a given n, we determine the maximal length of the file that the user can store, and thus quantify the necessary cost to store a file with a certain length, in terms of the length of the secret that the user needs to share with the servers. Additionally, for this maximal file length, we determine (i) the optimal amount of local randomness needed at the user, (ii) the optimal amount of public communication from the user to the servers, and (iii) the optimal amount of storage requirement at the servers.**

I. Introduction

Centralized data storage of sensitive information could mean compromising the entirety of the data in the case of a data breach. By contrast, a decentralized storage strategy can offer resilience against data breaches and avoid having a single point of entry for hackers. Well-known decentralized strategies are able to ensure that if a file is stored in L different servers, then any $t \leqslant L$ servers that pool their information can reconstruct the file, and $z < t$ compromised servers do not leak any information about the file in an information-theoretic sense. For instance, secret sharing [1], [2] solves this problem with the optimal storage size requirement at each server. Specifically, to store F bits over L servers, the best possible storage strategy, that allows reconstruction of the information from $t \leqslant L$ servers and is resilient against data breaches at $z < t$ servers, requires to store F bits of information in each of the L servers. In secret sharing models, the user who wishes to store a file in the servers corresponds to the dealer, the file corresponds to a secret, and the information stored at a given server is called a share of the secret. Applications of secret sharing to secure distributed storage have been extensively studied for a wide range of settings, e.g., [3]–[11]. Note that, as motivated in [12]–[15], the servers could also correspond to independent cloud storage providers, as it is often less costly for businesses and organizations to outsource data storage but cloud storage providers lack solid security guarantees and may be the victims of data breaches.

This work is supported in part by NSF grant CCF-2047913.

Because the user and the servers are not physically collocated, a standard assumption in secret sharing models [1], [2], [16]–[19] is that there exist individual and information-theoretically secure channels between the user and each server, in order to allow the user to give a share of the secret to the server. In this paper, we propose to quantify the cost associated with this assumption. Specifically, instead of assuming the availability at no cost of such information-theoretically secure channels, we assume that the user can communicate over a one-way public channel with each server, and that the user and each server share a secret key, which is a sequence of n bits uniformly distributed over $\{0,1\}^n$. Then, for a given n, we determine the maximal length of the file that the user can store. Given this relationship between n and the maximal length of the file, one can thus, for a given file length, determine the necessary cost to store a file, in terms of the secret length that the user needs to share with the servers. Furthermore, for a given secret key length n and the associated maximal length of the file that can be stored, we determine (i) the optimal amount of local randomness needed at the user, (ii) the optimal amount of public communication from the user to the servers, and (iii) the optimal amount of storage requirement at the servers.

The most challenging part of our work is to show the converse results on the maximal length of the file that the user can store, the optimal amount of local randomness needed at the user, and the optimal amount of public communication between the user and the servers. Unlike in traditional secret sharing models, in our converse, we need to account for the presence of shared secret keys, public communication available to all parties, and the fact that the creation phase of the shares and the distribution phase of the shares is allowed to be jointly performed in our model. Note that these two phases are completely independent in traditional secret sharing models, which only focus on the creation phase of the shares as the distribution phase of the shares relies on the assumption that information-theoretically secure channels are available at no cost. Finally, our achievability results, that match our converse results, can be obtained from ramp secret sharing schemes [20], [21].

In Section II, we formally state the problem. In Section III,

we present our main results. Finally, in Section IV, we provide 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 a given set $\mathcal{S}$, let $2^{\mathcal{S}}$ denote the power set of $\mathcal{S}$. Finally, let $\bigtimes$ 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 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* $\left(2^{r^{(F)}}, 2^{r^{(R)}}, \left(2^{r_l^{(M)}}\right)_{l\in\mathcal{L}}, \left(2^{r_l^{(S)}}\right)_{l\in\mathcal{L}}\right)$ *private file 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 the 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 functions $f_{\mathcal{A}} : \bigtimes_{l\in\mathcal{A}} \mathcal{S}_l \to \mathcal{F}$, 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 $\widehat{F}(\mathcal{A}) \triangleq f_{\mathcal{A}}(S_{\mathcal{A}})$, an estimate of F, where $S_{\mathcal{A}} \triangleq (S_l)_{l\in\mathcal{A}}$.*

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

$$\forall \mathcal{A} \subseteq \mathcal{L}, |\mathcal{A}| \geqslant t \implies H(F|\widehat{F}(\mathcal{A})) = 0 \text{ (Reliability)}, \quad (1)$$

$$\forall \mathcal{U} \subseteq \mathcal{L}, |\mathcal{U}| \leqslant z \implies I(F; M, K_{\mathcal{U}}) = 0 \text{ (Security)}. \quad (2)$$

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

(1) means that any subset of servers with size larger or equal than t is able to perfectly recover the files F, and (2) means that any subset of servers with size smaller or equal than z is unable to learn any information about the file.

Our main objective is to determine, under the constraints (1) and (2), 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 $\mathcal{C}_F(t,z)$. To this end, we introduce the following definition.

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

$$r_\star^{(F)}(t,z) \triangleq \sup_{r^{(F)}\in\mathcal{C}_F(t,z)} r^{(F)}, \quad (3)$$

$$r_{l,\star}^{(M)}(t,z) \triangleq \inf_{T\in\mathcal{Q}(r_\star^{(F)}(t,z))} r_l^{(M)}, l \in \mathcal{L}, \quad (4)$$

$$r_{\Sigma,\star}^{(M)}(t,z) \triangleq \inf_{T\in\mathcal{Q}(r_\star^{(F)}(t,z))} \sum_{l\in\mathcal{L}} r_l^{(M)}, \quad (5)$$

$$r_\star^{(R)}(t,z) \triangleq \inf_{T\in\mathcal{Q}(r_\star^{(F)}(t,z))} r^{(R)}, \quad (6)$$

$$r_{l,\star}^{(S)}(t,z) \triangleq \inf_{T\in\mathcal{Q}(r_\star^{(F)}(t,z))} r_l^{(S)}, l \in \mathcal{L}. \quad (7)$$

$r_\star^{(F)}(t,z)$ is the largest file size that the user can privately store under the constraints (1) and (2). Then, $r_\star^{(R)}(t,z)$, $r_{l,\star}^{(M)}(t,z)$, $r_{\Sigma,\star}^{(M)}(t,z)$, and $r_{l,\star}^{(S)}(t,z)$, $l \in \mathcal{L}$, 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 $r_\star^{(F)}(t,z)$. A priori, it is unclear whether there exists a $\left(2^{r_\star^{(F)}(t,z)}, 2^{r_\star^{(R)}(t,z)}, \left(2^{r_{l,\star}^{(M)}(t,z)}\right)_{l\in\mathcal{L}}, \left(2^{r_{l,\star}^{(S)}(t,z)}\right)_{l\in\mathcal{L}}\right)$ file storage strategy that (t,z)-achieves $r_\star^{(F)}(t,z)$.

III. Main results

A. Impossibility results

Theorem 1 (Converse on the file' length)**.** *Let $t \in [\![1,L]\!]$ and $z \in [\![1,t-1]\!]$. Then, we have*

$$r_\star^{(F)}(t,z) \leqslant n(t-z).$$

Proof. See Appendix A. ∎

Theorem 1 means that it is impossible for the user to store a file of length larger than $n(t-z)$ bits.

Theorem 2 (Converse on storage size requirement at the servers)**.** *Let $t \in [\![1,L]\!]$ and $z \in [\![1,t-1]\!]$. Then, we have*

$$r_{l,\star}^{(S)}(t,z) \geqslant n, \forall l \in \mathcal{L}.$$

Proof. See Appendix B. ∎

Theorem 2 means that in our setting 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)**.** *Let $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. Then, we have*

$$r^{(M)}_{\Sigma,\star}(t,z) \geqslant \frac{L}{t-z} r^{(F)}_{\star}(t,z). \tag{8}$$

Proof. See Appendix C. ∎

Theorem 3 means that it is impossible for the user to store a file of length $r^{(F)}_{\star}(t,z)$ if the public communication sum-length to the servers is smaller than $\frac{L}{t-z} r^{(F)}_{\star}(t,z)$ bits.

Theorem 4 (Converse on the amount of public communication to an individual server)**.** *Let $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. Consider the following condition*

$$\forall \mathcal{U}, \mathcal{V} \subseteq \mathcal{L}, |\mathcal{U}| = |\mathcal{V}| \Longrightarrow I\left(F; M_{\mathcal{U}}, K_{\mathcal{U}}\right) = I\left(F; M_{\mathcal{V}}, K_{\mathcal{V}}\right). \tag{9}$$

(9) *indicates that any two sets of colluding servers that have the same size have the same amount of information about the file F. If* (9) *holds, then we have*

$$r^{(M)}_{l,\star}(t,z) \geqslant \frac{1}{t-z} r^{(F)}_{\star}(t,z), \forall l \in \mathcal{L}.$$

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 $r^{(F)}_{\star}(t,z)$ if the public communication length to Server $l \in \mathcal{L}$ is smaller than $\frac{1}{t-z} r^{(F)}_{\star}(t,z)$ bits.

Theorem 5 (Converse on the amount of required local randomness at the user)**.** *Let $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. Then, we have*

$$r^{(R)}_{\star}(t,z) \geqslant \frac{z}{t-z} r^{(F)}_{\star}(t,z). \tag{10}$$

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

B. Capacity results

Theorem 6. *Let $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. We have*

$$\begin{aligned} r^{(F)}_{\star}(t,z) &= n(t-z), \\ r^{(R)}_{\star}(t,z) &= nz, \\ r^{(S)}_{l,\star}(t,z) &= n, \forall l \in \mathcal{L}, \\ r^{(M)}_{\Sigma,\star}(t,z) &= Ln, \\ (9) \implies \left(r^{(M)}_{l,\star}(t,z) = n, \forall l \in \mathcal{L}\right). & \end{aligned}$$

Theorem 6 provides a characterization of the quantities introduced in Definition 3.

Theorem 7. *Let $t \in [\![1, L]\!]$ and $z \in [\![1, t-1]\!]$. There exists a $\left(2^{r^{(F)}}, 2^{r^{(R)}}, \left(2^{r_l^{(M)}}\right)_{l\in\mathcal{L}}, \left(2^{r_l^{(S)}}\right)_{l\in\mathcal{L}}\right)$ private file storage strategy that (t,z)-achieves $r^{(F)}$ such that*

$$\begin{aligned} r^{(F)} &= r^{(F)}_{\star}(t,z), \\ r^{(R)}_{d} &= r^{(R)}_{\star}(t,z), \\ r^{(S)}_{l} &= r^{(S)}_{l,\star}(t,z), \forall l \in \mathcal{L}, \\ \textstyle\sum_{l\in\mathcal{L}} r^{(M)}_{l} &= r^{(M)}_{\Sigma,\star}(t,z), \\ r^{(M)}_{l} &= r^{(M)}_{l,\star}(t,z), \forall l \in \mathcal{L}, \text{ when (9) holds.} \end{aligned}$$

Theorem 7 shows that the optimal quantities of Definition 3 can be obtained simultaneously by a single private file storage strategy. The achievability part of Theorem 7 is obtained via ramp secret sharing schemes [20], [21] and is omitted.

IV. Concluding remarks

We considered the problem of storing a file in L servers such that any $t \leqslant L$ servers can reconstruct the file, and any subset of $z < t$ colluding servers cannot learn any information about the file. Unlike solutions that rely on traditional secret sharing models, we developed a new model that does not make the assumption that individual and information-theoretically secure channels between the user and each server are available at no cost. Instead, we assume that the user can communicate with the servers over one-way public channels, and share with each server a secret key with length n, which is meant to quantify the cost of privately storing the file. For a given secret-key length n and parameters t and z, we established the maximal length of the file that the user can store. Additionally, we determine in this case the minimum amount of local randomness needed at the user, the minimum amount of public communication between the user and the servers, and the minimum amount of storage space required at the servers.

Appendix A
Proof of Theorem 1

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

$$\begin{aligned} r^{(F)} &\overset{(a)}{=} H(F) \\ &= H(F|M, K_{\mathcal{U}}) + I(F; M, K_{\mathcal{U}}) \\ &\overset{(b)}{=} H(F|M, K_{\mathcal{U}}) \\ &= I(\widehat{F}(\mathcal{A}); F|M, K_{\mathcal{U}}) + H(F|M, K_{\mathcal{U}}, \widehat{F}(\mathcal{A})) \\ &\overset{(c)}{\leqslant} I(\widehat{F}(\mathcal{A}); F|M, K_{\mathcal{U}}) + H(F|\widehat{F}(\mathcal{A})) \\ &\overset{(d)}{=} I(\widehat{F}(\mathcal{A}); F|M, K_{\mathcal{U}}) \\ &\overset{(e)}{\leqslant} I(M, K_{\mathcal{A}}; F|M, K_{\mathcal{U}}) \\ &\overset{(f)}{=} I(K_{\mathcal{A}}; F|M, K_{\mathcal{U}}) \\ &\overset{(g)}{\leqslant} I(K_{\mathcal{A}}; K_{\mathcal{L}}, F, R|K_{\mathcal{U}}) \\ &= I(K_{\mathcal{A}}; K_{\mathcal{L}}, F|K_{\mathcal{U}}) + I(K_{\mathcal{A}}; R|K_{\mathcal{L}}, F) \\ &\overset{(h)}{=} I(K_{\mathcal{A}}; K_{\mathcal{L}}, F|K_{\mathcal{U}}) \end{aligned}$$

$$\begin{aligned}
&\overset{(i)}{=} I(K_{\mathcal{A}}; K_{\mathcal{L}}|K_{\mathcal{U}})\\
&\overset{(j)}{=} H(K_{\mathcal{A}}|K_{\mathcal{U}})\\
&\overset{(k)}{=} H(K_{\mathcal{A}\backslash\mathcal{U}})\\
&\overset{(l)}{=} n(t-z),
\end{aligned} \tag{11}$$

where (a) holds by uniformity of the file, (b) holds by (2) because $|\mathcal{U}|= z$, (c) holds because conditioning reduces entropy, (d) holds by (1) because $|\mathcal{A}|= t$, (e) holds because $\widehat{F}(\mathcal{A})$ is a function of $S_{\mathcal{A}}$ which is itself a function of $(M, K_{\mathcal{A}})$, (f) holds because $I(M; F|M, K_{\mathcal{U}}, K_{\mathcal{A}}) = 0$, (g) holds M is a function of $(F, K_{\mathcal{L}}, R)$, (h) holds by independence between R and $(K_{\mathcal{L}}, F)$, (i) holds by independence between F and $K_{\mathcal{L}}$, (j) holds because $\mathcal{A} \subseteq \mathcal{L}$, (k) holds because $\mathcal{U} \subset \mathcal{A}$, (l) holds because the keys K_l, $l \in \mathcal{A}\backslash\mathcal{U}$, are independent and each uniformly distributed over $\{0,1\}^n$ and $|\mathcal{A}\backslash\mathcal{U}|= t-z$.

Appendix B
Proof of Theorem 2

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

Appendix C
Proof of Theorem 3

For $\mathcal{T} \subseteq \mathcal{L}$ and $\mathcal{S} \subseteq \mathcal{L}\backslash\mathcal{T}$ such that $|\mathcal{T}|= z$ and $|\mathcal{S}|= t-z$, one can show that

$$\sum_{l\in\mathcal{S}} H(M_l) + \sum_{l\in\mathcal{S}} H(K_l) \geqslant H(F) + \sum_{l\in\mathcal{S}} H(K_l). \tag{12}$$

Next, we have

$$\begin{aligned}
&\frac{L}{t-z} r^{(F)}\\
&\overset{(a)}{=} \frac{L}{t-z} H(F)\\
&= \frac{L}{t-z}\binom{L}{z}^{-1}\binom{L-z}{t-z}^{-1} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=z}} \sum_{\substack{\mathcal{S}\subseteq\mathcal{T}^c\\|\mathcal{S}|=t-z}} H(F)\\
&\overset{(b)}{\leqslant} \frac{L}{t-z}\binom{L}{z}^{-1}\binom{L-z}{t-z}^{-1} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=z}} \sum_{\substack{\mathcal{S}\subseteq\mathcal{T}^c\\|\mathcal{S}|=t-z}} \sum_{l\in\mathcal{S}} H(M_l)\\
&\overset{(c)}{=} \frac{L}{t-z}\binom{L}{z}^{-1}\binom{L-z}{t-z}^{-1} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=z}} \binom{L-z-1}{t-z-1} \sum_{l\in\mathcal{T}^c} H(M_l)\\
&\overset{(d)}{=} \frac{L}{t-z}\binom{L}{z}^{-1}\binom{L-z}{t-z}^{-1}\binom{L-z-1}{t-z-1} \sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=L-z}} \sum_{l\in\mathcal{T}} H(M_l)\\
&\overset{(e)}{=} \frac{L}{t-z}\binom{L}{z}^{-1}\binom{L-z}{t-z}^{-1}\binom{L-z-1}{t-z-1}\binom{L-1}{L-z-1}\\
&\quad\times \sum_{l\in\mathcal{L}} H(M_l)\\
&= \sum_{l\in\mathcal{L}} H(M_l)\\
&\overset{(f)}{\leqslant} \sum_{l\in\mathcal{L}} r^{(M)}_l,
\end{aligned} \tag{13}$$

where (a) holds by uniformity of F, (b) holds by (12), (c) holds because for any $l \in \mathcal{T}^c$, $H(M_l)$ appears exactly $\binom{L-z-1}{t-z-1}$ times in the term $\sum_{\substack{\mathcal{S}\subseteq\mathcal{T}^c\\|\mathcal{S}|=t-z}} \sum_{l\in\mathcal{S}} H(M_l)$ (note that this observation was also made in [23, Lemma 3.2]), (d) holds by a change of variables in the sums, (e) holds because for any $l \in \mathcal{L}$, $H(M_l)$ appears exactly $\binom{L-1}{L-z-1}$ times in the term $\sum_{\substack{\mathcal{T}\subseteq\mathcal{L}\\|\mathcal{T}|=L-z}} \sum_{l\in\mathcal{T}} H(M_l)$, (f) holds by definition of M_l, $l \in \mathcal{L}$.

Appendix D
Proof of Theorem 4

Assume that (9) holds. Fix $l \in \mathcal{L}$. For $i \in [\![z, t-1]\!]$, define $\mathcal{V}_i \triangleq \begin{cases} [\![1,i]\!] & \text{if } l > i\\ [\![1,i+1]\!]\backslash\{l\} & \text{if } l \leqslant i\end{cases}$ and $\mathcal{V}_t \triangleq \mathcal{V}_{t-1}\cup\{l\}$. For $i \in \mathcal{L}$, and $\mathcal{S} \subseteq \mathcal{L}$ such that $|\mathcal{S}|= i$, define $\alpha_i \triangleq I\left(F; M_{\mathcal{S}}, K_{\mathcal{S}}\right)$ and $\alpha_{L+1} \triangleq \alpha_L$. Note that α_i only depends on i and not on $\mathcal{S}$ by (9). Note also that $\alpha_z = 0$ by (2) and $\alpha_t = H(F)$ by (1). Next, we have

$$\begin{aligned}
&H(M_l) + H(K_l)\\
&\geqslant H(M_l, K_l)\\
&\overset{(a)}{\geqslant} H(M_l, K_l|M_{\mathcal{V}_z}, K_{\mathcal{V}_z})\\
&\overset{(b)}{=} \sum_{i=z}^{t-1}\left[H(M_l, K_l|M_{\mathcal{V}_i}, K_{\mathcal{V}_i}) - H(M_l, K_l|M_{\mathcal{V}_{i+1}}, K_{\mathcal{V}_{i+1}})\right]
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\overset{(c)}{=} \sum_{i=z}^{t-1}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2} + H(M_l, K_l|F, M_{\mathcal{V}_i}, K_{\mathcal{V}_i})\\
&\quad - H(M_l, K_l|F, M_{\mathcal{V}_{i+1}}, K_{\mathcal{V}_{i+1}})]\\
&\overset{(d)}{\geqslant} [2\alpha_t - \alpha_{t-1} - \alpha_{t+1} + H(M_l, K_l|F, M_{\mathcal{V}_{t-1}}, K_{\mathcal{V}_{t-1}})]\\
&\quad + \sum_{i=z}^{t-2}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]\\
&\overset{(e)}{\geqslant} [2\alpha_t - \alpha_{t-1} - \alpha_{t+1} + H(M_l, K_l|F, M_{\mathcal{V}_{t-1}}, K_{\mathcal{V}_{t-1}})]^+\\
&\quad + \sum_{i=z}^{t-2}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+\\
&\overset{(f)}{=} [\alpha_t - \alpha_{t-1} + H(M_l, K_l|F, M_{\mathcal{V}_{t-1}}, K_{\mathcal{V}_{t-1}})]^+\\
&\quad + \sum_{i=z}^{t-2}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+\\
&\overset{(g)}{=} [\alpha_t - \alpha_{t-1} + H(M_l, K_l|F, M_{\mathcal{V}_{t-1}}, K_{\mathcal{V}_{t-1}})]\\
&\quad + \sum_{i=z}^{t-2}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+\\
&\overset{(h)}{=} H(M_l, K_l|F, M_{\mathcal{V}_{t-1}}, K_{\mathcal{V}_{t-1}}) + \sum_{i=z}^{t-1}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+
\end{aligned}$$

$$\begin{aligned}
&\overset{(i)}{\geqslant} H(K_l|F, M_{\mathcal{V}_{t-1}}, K_{\mathcal{V}_{t-1}}) + \sum_{i=z}^{t-1}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+ \\
&\overset{(j)}{\geqslant} H(K_l|F, R, K_{\mathcal{V}_{t-1}}) + \sum_{i=z}^{t-1}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+ \\
&\overset{(k)}{\geqslant} H(K_l) + \sum_{i=z}^{t-1}[2\alpha_{i+1} - \alpha_i - \alpha_{i+2}]^+ \\
&\overset{(l)}{\geqslant} H(K_l) + \min_{f\in\mathcal{F}} \sum_{i=1}^{t-z}[2f(i+1) - f(i) - f(i+2)]^+, \quad (15)
\end{aligned}$$

where (a) holds because conditioning reduces entropy, (b) holds because $l \in \mathcal{V}_t$, (c) holds by the chain rule and the definition of α_i, (d) holds because for any $i \in [\![z, t-2]\!]$, $H(M_l, K_l|F, M_{\mathcal{V}_i}, K_{\mathcal{V}_i}) \geqslant H(M_l, K_l|F, M_{\mathcal{V}_{i+1}}, K_{\mathcal{V}_{i+1}})$ since conditioning reduces entropy and $\mathcal{V}_i \subset \mathcal{V}_{i+1}$, and because $H(M_l, K_l|F, M_{\mathcal{V}_t}, K_{\mathcal{V}_t}) = 0$ since $l \in \mathcal{V}_t$, (e) holds because in (14), we observe that $H(M_l, K_l|M_{\mathcal{V}_i}, K_{\mathcal{V}_i}) - H(M_l, K_l|M_{\mathcal{V}_{i+1}}, K_{\mathcal{V}_{i+1}}) \geqslant 0$ since conditioning reduces entropy and $\mathcal{V}_i \subset \mathcal{V}_{i+1}$, (f) holds because $\alpha_{t+1} = \alpha_t = H(F)$ by (1), (g) holds because $\alpha_t \geqslant \alpha_{t-1}$ by the definition of α_t and α_{t-1}, (h) holds because $\alpha_t - \alpha_{t-1} = [2\alpha_t - \alpha_{t-1} - \alpha_{t+1}]^+$, (i) holds by the chain rule and non-negativity of the entropy, (j) holds because $M_{\mathcal{V}_{t-1}}$ is a function of $(F, R, K_{\mathcal{V}_{t-1}})$, (k) holds by independence between K_l and $(F, R, K_{\mathcal{V}_{t-1}})$ since $\{l\} \cap \mathcal{V}_{t-1} = \emptyset$, in (l) the minimum is taken over the set $\mathcal{F}$ of all the functions $f : [\![1, t-z+2]\!] \to [0, 1]$ that are non-decreasing (because, by construction, $(\alpha_i)_{i\in[\![1,L+1]\!]}$ is a non-decreasing sequence) and such that $f(1) = \alpha_z = 0$, $f(t-z+2) = f(t-z+1) = \alpha_t = H(F)$.

We now lower bound the minimum in the right-hand side of (15). Let $f \in \mathcal{F}$ and let f^+ be the concave envelope of f over $[\![1, t-z+2]\!]$, i.e., for $i \in [\![1, t-z+2]\!]$, $f^+(i) \triangleq \min\{g(i) : g \geqslant f, g \text{ is concave}\}$. Note that $f^+(1) = f(1)$ and $f^+(t-z+2) = f(t-z+2)$. Then, for any $i \in [\![1, t-z]\!]$ such that $f(i+1) = f^+(i+1)$, we have

$$\begin{aligned}
&[2f(i+1) - f(i) - f(i+2)]^+ \\
&\geqslant 2f(i+1) - f(i) - f(i+2) \\
&\overset{(a)}{\geqslant} 2f(i+1) - f^+(i) - f^+(i+2) \\
&\overset{(b)}{=} 2f^+(i+1) - f^+(i) - f^+(i+2), \quad (16)
\end{aligned}$$

where (a) holds because $f^+ \geqslant f$, (b) holds because $f(i+1) = f^+(i+1)$. Moreover, for any $i \in [\![1, t-z]\!]$ such that $f(i+1) \neq f^+(i+1)$, we have

$$\begin{aligned}
&[2f(i+1) - f(i) - f(i+2)]^+ \\
&\geqslant 0 \\
&= 2f^+(i+1) - f^+(i) - f^+(i+2), \quad (17)
\end{aligned}$$

where the last equality holds because f^+ is linear between i and $i+2$, i.e., $f^+(i+1) - f^+(i) = f^+(i+2) - f^+(i+1)$. Indeed, by contradiction, assume that f^+ is not linear between i and $i+2$, then we must have that

$$f^+(i+1) > \frac{f^+(i+2) + f^+(i)}{2} \quad (18)$$

since f^+ is concave. Next, we have a contradiction by constructing $h_i : [\![1, t-z+2]\!] \to \mathbb{N}$, a concave function such that $f \leqslant h_i < f^+$, as follows:

$$h_i : j \mapsto \begin{cases} f^+(j) & \text{if } j \neq i+1 \\ \max\left(\frac{f^+(i+2)+f^+(i)}{2}, f(i+1)\right) & \text{if } j = i+1 \end{cases}.$$

We have $f \leqslant h_i$ (since $f \leqslant f^+$), and $h_i < f^+$ by (18) and because $f^+(i+1) > f(i+1)$ (since $f^+ \geqslant f$ and $f^+(i+1) \neq f(i+1)$). Then, to show concavity of h_i, it is sufficient to show that h_i^Δ is non-increasing where h_i^Δ is defined as

$$\begin{aligned}
h_i^\Delta : [\![1, t-z+1]\!] &\to \mathbb{N} \\
j &\mapsto h_i(j+1) - h_i(j).
\end{aligned}$$

For $j \in [\![1, i-2]\!] \cup [\![i+2, t-z+1]\!]$, we have

$$h_i^\Delta(j+1) \leqslant h_i^\Delta(j) \quad (19)$$

by definition of h_i^Δ and concavity of f^+. Then, one can check that

$$\begin{aligned}
h_i^\Delta(i) &\leqslant h_i^\Delta(i-1), \\
h_i^\Delta(i+1) &\leqslant h_i^\Delta(i), \\
h_i^\Delta(i+2) &\leqslant h_i^\Delta(i+1).
\end{aligned}$$

Hence, h_i^Δ is non-increasing and we have proved (17) by contradiction. Next, we have

$$\begin{aligned}
&\sum_{i=1}^{t-z}[2f(i+1) - f(i) - f(i+2)]^+ \\
&\overset{(a)}{\geqslant} \sum_{i=1}^{t-z}[2f^+(i+1) - f^+(i) - f^+(i+2)] \\
&= \sum_{i=1}^{t-z}[(f^+(i+1) - f^+(i)) - (f^+(i+2) - f^+(i+1))] \\
&= f^+(2) - f^+(1) + f^+(t-z+2) - f^+(t-z+1) \\
&\overset{(b)}{=} f^+(2) \\
&\overset{(c)}{\geqslant} H(F)\frac{1}{t-z}, \quad (20)
\end{aligned}$$

where (a) holds by (16) and (17), (b) holds because $f^+(t-z+2) = f^+(t-z+1) = f(t-z+1) = H(F)$ and $f^+(1) = 0$, (c) holds because $f^+(2) = f^+(2) - f^+(1) \geqslant (f^+(t-z+1) - f^+(1))/(t-z)$ by concavity of f^+ and where we have used that $f^+(t-z+1) = H(F)$ and $f^+(1) = f(1) = 0$. Finally, we have

$$r_l^{(M)} \geqslant H(M_l) \overset{(a)}{\geqslant} H(F)\frac{1}{t-z} \overset{(b)}{=} r^{(F)}\frac{1}{t-z}, \quad (21)$$

where (a) holds by (15) and (20), which is valid for any $f \in \mathcal{F}$, (b) holds by uniformity of F.

REFERENCES

[1] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

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

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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)*. IEEE, 2020, pp. 605–610.

[12] A. Bessani, M. Correia, B. Quaresma, F. André, 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.

[13] 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.

[14] 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.

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

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

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

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] A. De Santis and B. Masucci, "Multiple ramp schemes," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1720–1728, 1999.