Security in Locally Repairable Storage

Abhishek Agarwal and Arya Mazumdar

***Abstract*—In this paper we extend the notion of *locally repairable codes* to *secret sharing* schemes. The main problem we consider is to find the optimal ways to distribute shares of a secret among a set of storage-nodes (participants) such that the content of each node (share) can be recovered by using contents of only few other nodes, and at the same time the secret can be reconstructed by only some allowable subsets of nodes. As a special case, an eavesdropper observing some set of a specified nodes (such as less than certain number of nodes) does not get any information. In other words, we propose to study a locally repairable distributed storage system that is secure against a *passive eavesdropper* that can observe some subsets of nodes.**

We provide a number of results related to such systems including upper-bounds and achievability results on the number of bits that can be securely stored with these constraints.

I. Introduction

We start with a bit of notation. Suppose $[n] \equiv \{1, 2, \ldots, n\}$. For any set A, 2^A denotes the power set of A. For any vector $\mathbf{x} \in \mathbb{F}_q^n$ and $I \subseteq [n]$, $\mathbf{x}_I$ denotes the projection of $\mathbf{x}$ onto the coordinates I. For singleton sets we write $x_i \equiv \mathbf{x}_{\{i\}}$. $a|b$ denotes a divides b. The unit of entropy in this paper is q-ary. An (n, k)-MDS code represents a maximum distance separable (MDS) code with length n and dimension k over some finite field.

Secret sharing schemes were proposed by Shamir and Blakley [2], [17] to provide security against an eavesdropper with infinite computational capacity. Consider the secret as a realization of a uniform random variable (rv) $\mathbf{s}$. Suppose that shares of the secret are to be distributed among n participants (storage nodes) such that a set of shares belonging to $\mathcal{A}_s \subset 2^{[n]}$, is able to determine the secret. $\mathcal{A}_s$ is called the access structure of the secret sharing scheme. Denote the share of a participant (or node) $i \in [n]$ by c_i and let $\mathbf{c} = (c_1 c_2 \ldots c_n)$. A secure scheme has the property that a subset of shares in the block-list $\mathcal{B}_s \subset 2^{[n]}$ are unable to determine anything about the secret. Thus, $H(\mathbf{s}|\mathbf{c}_B) = H(\mathbf{s})$ for any $B \in \mathcal{B}_s$ and $H(\mathbf{s}|\mathbf{c}_A) = 0$ for any $A \in \mathcal{A}_s$, where $H(\cdot)$ denotes the entropy. For a standard *monotone* secret sharing scheme the classes $\mathcal{A}_s$ and $\mathcal{B}_s$ must have the following properties, $A' \supset A, A \in \mathcal{A}_s \implies A' \in \mathcal{A}_s; B' \subset B, B \in \mathcal{B}_s \implies B' \in \mathcal{B}_s$; and $\mathcal{B}_s \subset 2^{[n]} \setminus \mathcal{A}_s$. For a *perfect* secret sharing scheme we have the above monotone property and $\mathcal{B}_s = 2^{[n]} \setminus \mathcal{A}_s$. Perfect schemes for access structures of the form $\mathcal{A}_s = \{A \subset [n] : |A| \geq m\}$ are called threshold secret sharing schemes. We refer to [1] for a comprehensive survey of secret sharing schemes.

A convenient property of schemes that need to store data in a distributed storage system is local repairability [6] i.e. any storage node can be repaired by accessing a small subset of other nodes, much smaller than is required for decoding the complete data. Error-correcting codes with the local repair property – locally repairable codes (LRC) – have been the center of a lot of research activities lately [3], [6], [11], [18]. Consider an n length code over a q-ary alphabet, $\mathcal{C} \subset \mathbb{F}_q^n$ of size $|\mathcal{C}| = q^k$. The code is said to have locality r, if for every i, $1 \leq i \leq n$, there exists a set $\mathcal{R}_i \subset [n] \setminus \{i\}$ with $|\mathcal{R}_i| \leq r$ such that for any two codewords $\mathbf{c}, \mathbf{c}'$ satisfying $c_i \neq c'_i$, we have $\mathbf{c}_{\mathcal{R}_i} \neq \mathbf{c}'_{\mathcal{R}_i}$. In a code with locality r, any symbol of a codeword can be deduced by reading only at most r other symbols of the codeword. Usually for the applications to distributed storage, the code is further required to have a *minimum distance* d. It is known that [6] for such a code,

$$d \leq n - k - \lceil k/r \rceil + 2, \tag{1}$$

which is also achievable [11], [18]. A q-ary code of length n, size q^k and locality r will be called an $(n, k, r)_q$ code if it's minimum distance satisfies (1) with equality.

Security in distributed storage has recently been considered in a number of papers, for example [7], [12], [15], [19] and references therein. In these papers the main objective is to secure stored data or repair data against an adversary. Threshold secret sharing protocols over a network under some communication constraint has been considered in [16]. Problems most closely related to this paper perhaps appear in [13] where a version of threshold secret sharing scheme with locality has been studied. Motivated by the above applications in distributed storage, we analyze secret sharing schemes with arbitrary access structures such that shares of each participant/node can be repaired with locality r.

A. Result and organization

In section II, we provide bounds and achievability results for a locally repairable secret sharing scheme that is most relevant to today's distributed storage systems. The particular access structure and block-list we consider are $\mathcal{A}_s = \{A \subset [n] : |A| \geq m\}$ and $\mathcal{B}_s = \{B \subset [n] : |B| \leq \ell\}$ respectively. We assume the shares of the secrets are locally recoverable and at the same time an adversary observing up to a number of shares do not get any information. This section also addresses the conditions under which a locally reparable code can be converted into a secret

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, email: {agarw050,arya}@umn.edu. This work was supported in part by NSF CCF 1318093 and CCF 1453121.

sharing scheme. We then extend the notion of security to co-operative repair where a Distributed Storage System (DSS) can deal with simultaneous multiple node failures. A different, more practical generalization for secret sharing scheme is made in which the DSS is represented by a graph $\mathcal{G}$ such that a node can only connect to its neighbors in $\mathcal{G}$ for repair. We provide upper-bounds on the secrecy capacity and construct achievable schemes in these scenario in sections III and IV.

We also consider perfect secret sharing schemes over general access structures under locality constraints. While we show that for threshold secret sharing schemes, there cannot exist any non-trivial locally repairability, we also give an example of a perfect secret sharing scheme with small locality. We show how locality effects the sizes of shares in a perfect scheme as they relate to the size of the secret. These results are presented in section V.

II. An imperfect scheme for secure storage

We start this section by formally defining a particular example of common access structure.

Definition 1. *An $(n, k, \ell, m, r)_q$-secret sharing scheme consists of a randomized encoder f that maps a uniform secret $\mathbf{s} \in \mathbb{F}_q^k$, to $\mathbf{c} = f(\mathbf{s}) \in \mathbb{F}_q^n$, and must have the following three properties.*

1) *(Recovery) Given any m symbols of $\mathbf{c}$, the secret $\mathbf{s}$ is completely determined. This guarantees that the secret is recoverable even with the loss of any $n - m$ shares.*
$$H(\mathbf{s}|\mathbf{c}_I) = 0,\ \forall I \subset [n], |I| = m \tag{2}$$
2) *(Security) Any set of ℓ shares of $\mathbf{c}$ does not reveal anything about the secret.*
$$H(\mathbf{s}|\mathbf{c}_J) = H(\mathbf{s}),\ \forall J \subset [n], |J| = \ell \tag{3}$$
A scheme satisfying this condition is called ℓ-secure.
3) *(Locality) For any share, there exist at most r other shares that completely determine this. For all i, there exists $\mathcal{R}_i \subset [n] \setminus \{i\} : |\mathcal{R}_i| \leq r$, such that*
$$H(c_i|\mathbf{c}_{\mathcal{R}_i}) = 0 \tag{4}$$
$\mathcal{R}_i$ is called the recovery set of share i.

The maximum amount of secret that can be stored as a function of n, ℓ, m and r is called the capacity of the secret sharing scheme and in the following we provide exact characterization of this quantity. We can define the security condition above in a modified way where the eavesdropper is allowed to see any set $J \subset [n]$ of shares and we calculate the amount of information revealed, i.e. $I(\mathbf{s}; \mathbf{c}_J)$, in terms of $n, k, |J|, m$ and r in an optimal scheme. This extension is evident from our result and somewhat summarized in corollary 2.

Note that, for locally repairable schemes with no security requirement i.e. $\ell = 0$ the following lower-bound on m is apparent from (1),
$$m \geq k + \lceil k/r \rceil - 1, \tag{5}$$
This lower bound follows from the definition of the minimum distance of a code $d = n - m + 1$. In the subsequent, we provide a fundamental limit and constructions achieving that limit.

A. Bounds

The converse result that can be obtained for the aforementioned access structure is following.

Theorem 1. *Any $(n, k, \ell, m, r)_q$-secret sharing scheme must satisfy,*
$$k + \ell \leq m - \left\lfloor \frac{m}{r+1} \right\rfloor. \tag{6}$$

To prove (6), using eq. (1) as a black-box will not work. Instead we can follow its proof method [3], [6]. The upper-bound in eq. (6) can also be obtained from [13, Thm. 33] where the authors use a different method. We omit the proof here. Indeed, we can show the following,

Corollary 2. *There exist sets $J \subset [n]$ with $\ell \leq |J| \leq m - \lfloor \frac{m}{r+1} \rfloor$ such that,*
$$H(\mathbf{s}|\mathbf{c}_J) \leq m - \left\lfloor \frac{m}{r+1} \right\rfloor - |J|. \tag{7}$$

Equation (7) gives an upper-bound on the maximum ambiguity of the secret of an (n, k, ℓ, m, r)-scheme when the eavesdropper has access to more than ℓ shares.

B. Constructions

It is possible to show matching achievability results to Theorem 1 by a number of different methods.

Theorem 3. *There exists a (n, k, ℓ, m, r)-secret sharing scheme such that eq. (6) is satisfied with equality.*

In particular this theorem can be proved by constructing a random linear network code. The proof will appear in the full version of this paper.

The achievability result also follows from [13], that gives a construction for optimal secure LRC employing Gabidulin codes to satisfy the security constraint. In the subsequent we outline their method that we adapt later for more general secret sharing schemes.

An intuitive construction of secure LRC codes comes by replacing some inputs to the LRC with uniform rv(s). Formally, consider a linear code $\mathcal{C}$ with code-length n and input size $(k + \ell)$. Let $G_{n\times(k+\ell)} = [G^1_{n\times\ell}\ G^2_{n\times k}]$ be the generator matrix of this code and $\mathbf{a} \in \mathbb{F}_q^{k+\ell}$ be the input to $\mathcal{C}$. Denote by $\mathbf{s} \in \mathbb{F}_q^k$ the input we want to store securely. We construct an ℓ secure secret sharing scheme using $\mathcal{C}$ by taking,
$$\mathbf{a} = [\mathbf{r}\ \mathbf{s}] \tag{8}$$
where $\mathbf{r} \in \mathbb{F}_q^\ell$ is a uniformly distributed random vector. This scheme is ℓ-secure iff for any ℓ *linearly independent* (LI) rows of G the corresponding rows of G^1 are LI.

Lemma 4. *Let $\mathbf{g}_i = [g_{i1} g_{i2} \dots g_{i(k+\ell)}], i \in [\ell]$ be any ℓ LI rows of G. The secret sharing scheme constructed in*

eq. (8) *is ℓ-secure iff the corresponding row vectors of* G^1, $\mathbf{g}_i^1 = [g_{i1}\, g_{i2} \dots g_{i\ell}], i \in [\ell]$ *are LI.*

We omit the proof of this lemma. Note that using lemma 4 we can add the security property to any linear code; we do not assume any locality property for the generator matrix G. But, it is clear that if the generator matrix G has locality r, then so would the scheme constructed in eq. (8). The construction of an optimal (n, k, l, m, r)−secret sharing scheme is described below.

Let $\alpha_1, \alpha_2 \in \mathbb{F}_{q^m}$. These points can be represented as vectors in $\mathbb{F}_q^m$. The points α_1, α_2 are said to be $\mathbb{F}_q$-*linearly independent* when the corresponding vectors in $\mathbb{F}_q$ are linearly independent. A Gabidulin code from $\mathbb{F}_{q^m}^k \to \mathbb{F}_{q^m}^n$, for input $(f_1 f_2 \dots f_k), f_i \in \mathbb{F}_{q^m}$, is obtained by evaluating the linearized polynomial $\Theta(y) = \sum_{i=1}^{k} f_i y^{q^{i-1}}$ at n $\mathbb{F}_q$-linearly independent points $\alpha_i \in \mathbb{F}_{q^m}, i \in [n]$. Note that we need $m \geq n$ to obtain n $\mathbb{F}_q$-linearly independent points in $\mathbb{F}_{q^m}$

Consider the generator matrix, $G_{n\times(k+\ell)} = [\mathbf{g}_1 \dots \mathbf{g}_n]^T$ of a linear $(n, k+\ell, r)_q$ code. Consider $\mathbf{a} = (\mathbf{s} \;\; \mathbf{r}), \mathbf{s} \in \mathbb{F}_{q^N}^k, \mathbf{r} \in \mathbb{F}_{q^N}^\ell$ with $N \geq n$, where $\mathbf{r}$ is a uniformly distributed rv and $\mathbf{s}$ denotes the secret. First, $\mathbf{a}$ is precoded using a Gabidulin code, $\Gamma : \mathbb{F}_{q^N}^{k+\ell} \to \mathbb{F}_{q^N}^{k+\ell}$ which is obtained by evaluating the polynomial,

$$\Psi_{\mathbf{a}}(y) = \sum_{i=1}^{k+\ell} a_i y^{q^{i-1}} \tag{9}$$

at the $\mathbb{F}_q$-linearly independent points $\alpha_i \in \mathbb{F}_{q^N}, i \in [k+\ell]$. Now, representing $\Gamma(\mathbf{a}) \in \mathbb{F}_{q^N}^{k+\ell}$ as a matrix of size $(k+\ell)\times N$ in $\mathbb{F}_q$ each column of the matrix can be encoded independently using the linear LRC to get $\mathbf{c} = (c_i)_{i=1}^n \in \mathbb{F}_{q^N}^i$. It is easy to show that this construction is ℓ-secure. The optimality of the scheme then follows from the optimality of the initial linear LRC. The proof of security of this construction is omitted.

C. Constructions with small alphabet size

Note that, the size of the alphabet in the construction of optimal secure LRC using Gabidulin codes is exponential in the number of nodes. Recently an optimal construction of locally repairable codes was proposed in [18] for general values of the parameters n, k, and r and alphabet size of $O(n)$. We use the construction in eq. (8) to form a secure scheme from the LRCs in [18] with a small alphabet and analyze the conditions for that construction to satisfy lemma 4. We assume that $(r+1)|n$ throughout this subsection.

We will need the following definition of Maximally Recoverable codes. Recall that, we denote by $(n, k, r)_q$ an optimal LRC with length n, size q^k, and locality r.

Definition 2. *Consider an* $(n, k, r)_q$*-LRC. Let* $\mathcal{Q}_j : |\mathcal{Q}_j| = r+1, j \in [n/(r+1)]$ *denote a partition of* $[n]$ *such that the recovery set is,*

$$\mathcal{R}_i = \mathcal{Q}(i) \setminus \{i\} \tag{10}$$

where $\mathcal{Q}(i) \in \{\mathcal{Q}_j\}_j$ *is the partition containing node* i*. Denote such an LRC by* $(n, k, r, \{\mathcal{Q}_j\}_j)_q$*. The* $(n, k, r, \{\mathcal{Q}_j\}_j)_q$*-LRC is called maximally recoverable [5] if the code obtained by puncturing one symbol from each* $\mathcal{Q}_j$ *is MDS.*

Note that [6] pointed out that an optimal LRC must have the recovery structure as in eq. (10).

The main objective of this section is to show that the immediate construction of (n, k, ℓ, m, r)-secret-sharing scheme from [18] is effective if and only if the code is maximally recoverable.

Denote the linear $(n, k, r, \{\mathcal{Q}_j\}_j)_q$-LRC proposed in [18] as $[n, k, r]_q - \text{TB}$ code. For input $\mathbf{a} \in \mathbb{F}_q^{k+\ell}$, the codeword for $[n, k, r]_q - \text{TB}$ code is obtained by evaluating the polynomial $f_{\mathbf{a}}(x) = \sum_{i=0}^{k+\ell-1} a_i x^{i \mod r} g(x)^{\lfloor i/r \rfloor}$ (where $g(\cdot)$ of degree $r+1$ is a polynomial that is constant on each of the partitions $\mathcal{Q}_j$) at points of $\mathcal{A} = \{\alpha_i\}_{i=1}^n \subset \mathbb{F}_q$. Let $G_{n\times(k+\ell)} = \left[\alpha_i^{(j-1) \mod r} g(\alpha_i)^{\lfloor (j-1)/r \rfloor}\right]_{i=1,j=1}^{n,k}$ be the generator matrix of the code and let $G = \left[G^1_{n\times\ell}\; G^2_{n\times k}\right]$. Note that G^1 represents the generator matrix for an $[n, \ell, r]_q - \text{TB}$ code with a smaller input size ℓ, but the same evaluation set $\mathcal{A}$ (code length n) and locality r.

Lemma 5. *Consider an* $[n, k+\ell, r]_q - TB$ *code with a generator matrix* $G_{n\times(k+\ell)}$*. Let* $\mathcal{S} \subset [n] : |\mathcal{S}| = \ell$ *and* $|\mathcal{S} \cap \mathcal{Q}_j| \leq r, j \in [n/(r+1)]$*. Then, the rows corresponding to* $\mathcal{S}$ *in* G *are linearly independent for any* $\ell \leq r - 1 + (r\lfloor k/(r-1)\rfloor - k)$

We omit the proof of this lemma. For $\ell < r$, the construction (in eq. (8)) using $[n, k+\ell, r]_q - \text{TB}$ code is ℓ-secure since any ℓ rows of G_1 form an $\ell \times \ell$ Vandermonde matrix. For $\ell > r$, we have the following result, using definition 2 and lemma 5.

Theorem 6. *Let* $g(x)$ *be a polynomial such that the* $[n, \ell, r]_q - TB$ *code constructed using* $g(x)$ *is maximally recoverable. Construct an* $[n, k+\ell, r]_q - TB$ *code using the same polynomial. The secret sharing scheme constructed in eq.* (8) *using this* $[n, k+\ell, r]_q - TB$ *code is ℓ-secure. Conversely, for a* $g(x)$ *generating an* $[n, \ell, r]_q - TB$ *code which is not maximally recoverable, the construction in eq.* (8) *is not ℓ-secure for* $\ell \in \left(r, r - 1 + (r\lfloor k/(r-1)\rfloor - k)\right]$.

Proof: Consider $\ell > r$. Let $g(x)$ be such that the $[n, \ell, r]_q - \text{TB}$ code is maximally recoverable. Let $G = [G^1_{n\times\ell} G^2_{n\times k}]$ be the generator matrix for the $[n, k+\ell, r]_q - \text{TB}$ constructed using the same good polynomial. Therefore, G^1 is the generator matrix for the $[n, \ell, r]_q - \text{TB}$ code. Consider a set $\mathcal{D} \subset [n]$ of any ℓ linearly dependent rows of G^1. Therefore, $\mathcal{Q}_j \subset \mathcal{D}$ for atleast one $j \in [n/(r+1)]$. Hence, the corresponding rows in G must also be linearly dependent. Thus, from lemma 4 the secret sharing construction in eq. (8) must be ℓ-secure.

Now, suppose that the $[n, \ell, r]_q - \text{TB}$ code is not maximally recoverable. Thus, there would exist an $\mathcal{S} \subset [n] : |\mathcal{S} \cap \mathcal{Q}_j| \leq r, \forall j \in [n/(r+1)]$ such that the rows in G^1 corresponding to $\mathcal{S}$ are linearly dependent. Now from lemma 5 we know that the rows corresponding to $\mathcal{S}$ in G are not linearly dependent for $r < \ell \leq r - 1 + (r\lfloor k/(r-1)\rfloor - k)$. Hence, from lemma 4 the secret sharing scheme cannot be ℓ secure. ∎

In the next two sections we extend the notions and results of this section to other generalized repair condi-

tions related to distributed storage.

III. Security for Schemes with co-operative repair

Co-operative repair for a locally repairable scheme addresses simultaneous multiple failures in a distributed storage system. To this end, we extend the definition in eq. (4) to a (r,δ) scheme where any δ –instead of just one– shares can be recovered from r other shares.

Definition 3. *A set $\mathcal{C} \subset \mathbb{F}_q^n$ is said to be (r,δ)-repairable if for every $\Delta \subset [n] : |\Delta| \leq \delta$ there exists a set $\mathcal{R}(\Delta) \subset [n] \setminus \Delta : |\mathcal{R}(\Delta)| \leq r$ such that for all $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$,*

$$\mathbf{c}_{1\Delta} \neq \mathbf{c}_{2\Delta} \implies \mathbf{c}_{1\mathcal{R}(\Delta)} \neq \mathbf{c}_{2\mathcal{R}(\Delta)} \tag{11}$$

Using definition 3 we can generalize the notion of an $(n,k,\ell,m,r)_q$-secret sharing scheme. For simplicity, in this case, we do not address the possibility of a catastrophic failure i.e. we take $m = n$. For this system we derive an upperbound on the capacity k given $n, \ell, r,$ and δ.

Definition 4. *An $(n,k,\ell,(r,\delta))_q$-secret sharing scheme consists of a randomized encoder $f(.)$ that stores the uniformly distributed file $\mathbf{s} \in \mathbb{F}_q^k$, in n separate shares such that the scheme is (r,δ)-repairable (definition 3) and ℓ-secure (cf. eq. (3)).*

Such schemes have been considered in [14] for $\ell = 0$ (no security) and the following upper-bound on the rate of such codes has been proposed,

$$R = \frac{k}{n} \leq \frac{r}{r+\delta}.$$

For the case of ℓ-secure codes we give an analogous upper bound on the rate of a secret sharing scheme in the following.

Theorem 7. *The rate $R = k/n$ of an $(n,k,l,(r,\delta))_q$ secret sharing scheme is bounded as,*

$$R \leq \frac{r}{r+\delta} - \frac{\ell}{n}. \tag{12}$$

It is easy to see that Gabidulin precoding (eq. (9)) would give an ℓ-secure construction with alphabet $\mathbb{F}_{q^N}$, for $N \geq n$, from any optimal linear $(n, k+\ell, 0, (r,\delta))_q$ code. Using the constructions of [14], it is therefore possible to obtain a rate of

$$\frac{k}{n} \geq \frac{r-\delta}{r+\delta} - \frac{\ell}{n}.$$

IV. Security for repairable codes on graphs

Consider a distributed storage system (DSS) as a directed graph $\mathcal{G}$ such that a node of the graph represents a node of the DSS and each node can connect to only its out-neighbors for repair [8], [9]. We define an ℓ-secure code in this scenario as follows.

Definition 5. *Let $\mathcal{G} = ([n], E)$ be a directed graph on n nodes. An $(n,k,\ell,m,\mathcal{G})_q$-secret sharing scheme consists of a randomized encoder that can store a uniform secret $\mathbf{s} \in \mathbb{F}_q^k$ on n shares/nodes, $\mathbf{c} = f(\mathbf{s}), \mathbf{c} \in \mathbb{F}_q^n$, such that the system is ℓ-secure (cf. eq. (3)) and the data can be recovered from any m shares (cf. eq. (2)). In addition the data on any node can be recovered form its neighbors i.e.*

$$H(\mathbf{c}_i | \mathbf{c}_{N(i)}) = 0$$

where $N(i)$ denotes the out-neighbours of node i in the graph $\mathcal{G} = ([n], E)$.

A bound on the capacity of such a scheme for $\ell = 0$ (no security) was derived in [10],

$$m \geq k + \max_{\substack{U \in \mathcal{I}(\mathcal{G}): \\ |N(U)| \leq k-1}} |U| \tag{13}$$

where $\mathcal{I}(\mathcal{G})$ denotes the induced acyclic graphs in $\mathcal{G}$, $N(U)$ the neighbours of U, and d is the minimum distance for the code. The lower bound on m for an ℓ-secure scheme on a graph $\mathcal{G}$ is given in the following.

Theorem 8. *For any $(n,k,\ell,m,\mathcal{G})_q$-secret sharing scheme on a graph $\mathcal{G}$, m satisfies the following lower bound,*

$$m \geq k + \ell + \max_{\substack{U \in \mathcal{I}(\mathcal{G}): \\ |N(U)| \leq \ell+k-1}} |U| \tag{14}$$

where $\mathcal{I}(G)$ denotes the set of induced acyclic graphs in $\mathcal{G}$.

We also note the following result.

Theorem 9. *Consider an $(n,k,\ell,n,\mathcal{G})_q$ secret sharing scheme. The secrecy capacity of the scheme satisfies the following upper-bound.*

$$k \leq n - |U| - |\ell| \tag{15}$$

where U is the largest acyclic induced subgraph in $\mathcal{G}$.

The proofs of the theorems are omitted. Constructions of schemes for secure repairable codes on graphs follows from the techniques of [10] and will appear at the full version of this paper.

V. Perfect Secret Sharing and General Access Structures

In this section we provide results regarding existence of locally repairable of perfect secret sharing schemes and the relation between sizes of shares and secret in those schemes.

A. Perfect access structures with locality

To make the (n,k,ℓ,m,r) secret sharing scheme perfect, we must have $m = \ell + 1$. This results in a threshold secret-sharing scheme. Now, from eq. (6) we have,

$$k \leq 1 - \left\lfloor \frac{\ell+1}{r+1} \right\rfloor.$$

Thus, for storing any secret we must have $r \geq \ell + 1 = m$. Since any secret sharing scheme works when $r \geq m$ (local repair in this case imply full revelation of secret) only

trivial locally repairable codes are possible for threshold secret sharing schemes. This imply the following statement.

Proposition 10. *A threshold secret sharing scheme is not locally repairable.*

We know that for perfect secret sharing schemes the size of the secret cannot be more than the size of any share [1, Lemma 2]. Thus, it seems that for locally repairable perfect secret sharing the secrecy capacity should be zero. We show that this is not true for general access structures. Indeed, the following is true.

Proposition 11. *There exists an access structure $\mathcal{A}_s$, for which a perfect secret sharing scheme is possible with non-trivial locality r i.e. $r < \min_{A\in\mathcal{A}_s}|A|$.*

Proof: Let n, κ be such that $r|\kappa$ and $(r+1)|n$. Consider an $(n, \kappa, r, \{\mathcal{Q}_j\}_j)$ maximally recoverable LRC (definition 2). We know that such codes exist from [5]. Now, we use the Gabidulin precoding method described above to construct a $(n, k = 1, \ell = \kappa - 1, m = \kappa(1 + 1/r), r)$ secret sharing scheme from this code.

Define the access structure to be $\mathcal{A}_s = \{A \subset [n] : \sum_{j=1}^{n/(r+1)} \min\{|A \cap \mathcal{Q}_j|, r\} \geq \kappa\}$. Now given any $A \in \mathcal{A}_s$, a user accessing the shares corresponding to A can determine the secret s_0 because the set always contains k shares of a punctured $(nr/(r+1), \kappa)$-MDS code.

For a perfect secret sharing scheme the block-list is given by $\mathcal{B}_s = \{B : \sum_{j=1}^{n/(r+1)} \min\{|B \cap \mathcal{Q}_j|, r\} < \kappa\}$. Assume that the eavesdropper has access to a set $B \in \mathcal{B}_s$. Construct the following set of size atmost $\kappa - 1$ from B,

$$B' = \cup_{j=1}^{n/(r+1)} B'_j, B' \subset B$$

where $B'_j \subset B_j, B_j = B \cap \mathcal{Q}_j$ is obtained by removing any one co-ordinate if $|B_j| > r$, otherwise $B'_j = B_j$. Note that $|B'| < \kappa$. Since all the shares in B are recoverable from $B' \subset B$, an eavesdropper with access to the nodes in B is equivalent to an eavesdropper with access to B'. And since $|B'| \leq \ell = \kappa - 1$, the eavesdropper does not get any information about the secret. ∎

B. Size of a share for perfect secret sharing with locality

We know that, for perfect secret sharing schemes, the size of the secret cannot be larger than the size of a share [1, Lemma 2]. In [4] the minimum node storage required for arbitrary monotone access structures is analyzed. It constructs an access structure for which the sizes of the shares has to be $n/\log(n)$ times the size of the secret for any perfect scheme. For secret sharing schemes with local repairability and fixed recovery sets, all monotone access structures are not feasible. The minimal sets of the access structure cannot include any recovery set. Here, we extend the result in [4] to the restricted class of monotone access structures.

Suppose that a secret file is stored on shares $[n]$ and the shares have locality r (eq. (4)). Consider a partition of $[n]$, $\mathcal{Q}_j : \mathcal{Q}_j, j \in [n/(r+1)]$ such that the recovery sets are given by eq. (10). For a perfect secret sharing scheme on $[n]$ with monotone access structure $\mathcal{A}_s$, the minimal sets of $\mathcal{A}_s$, $\mathcal{A}^\star_s$, must satisfy,

$$A \in \mathcal{A}^\star_s \implies A \not\supset \mathcal{Q}_j \ . \tag{16}$$

Denote this class of monotone access structures with $\mathbb{A}_s$. We have the following result for the minimum size of a share for secret sharing schemes with access structure $\mathcal{A}_s \in \mathbb{A}_s$.

Theorem 12. *Consider distribution of shares to n nodes with locality r, recovery sets as in eq.* (10)*. Then, there is an access structure $\mathcal{A}_s \in \mathbb{A}_s$ (eq.* (16)*), such that any perfect scheme, if exists, must contain a share of size at least $\frac{(r+1)n}{r \log n}$ times the size of the secret.*

We omit the proof of this theorem.

REFERENCES

[1] A. Beimel. Secret-sharing schemes: A survey, 2011.

[2] G. R. Blakley. Safeguarding cryptographic keys. In *Managing Requirements Knowledge, International Workshop on*, pages 313–313. IEEE Computer Society, 1899.

[3] V. Cadambe and A. Mazumdar. An upper bound on the size of locally recoverable codes. In *Proc. IEEE Int. Symp. Network Coding*, June 2013.

[4] L. Csirmaz. The Size of a Share Must Be Large. *Journal of Cryptology*, 10(4):223–231, Nov. 1997.

[5] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin. Explicit maximally recoverable codes with locality. *Computing Research Repository*, abs/1307.4150, 2013.

[6] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. *IEEE Trans. Inform. Theory*, 58(11):6925–6934, Nov. 2012.

[7] S. Goparaju, S. El Rouayheb, R. Calderbank, and H. V. Poor. Data secrecy in distributed storage systems under exact repair. In *Network Coding (NetCod), 2013 International Symposium on*, pages 1–6. IEEE, 2013.

[8] A. Mazumdar. Achievable schemes and limits for local recovery on a graph. In *Proc. Allerton Conf. Commun., Contr., Computing*, 2014.

[9] A. Mazumdar. On a duality between recoverable distributed storage and index coding. In *Proc. Int. Symp. Inform. Theory*, pages 1977–1981. IEEE, 2014.

[10] A. Mazumdar. Storage capacity of repairable networks. *arXiv preprint arXiv:1408.4862*, 2014.

[11] D. S. Papailiopoulos and A. G. Dimakis. Locally repairable codes. In *Proc. Int. Symp. Inform. Theory*, pages 2771–2775, Cambridge, MA, July 2012.

[12] S. Pawar, S. El Rouayheb, and K. Ramchandran. Securing dynamic distributed storage systems against eavesdropping and adversarial attacks. *Information Theory, IEEE Transactions on*, 57(10):6734–6753, 2011.

[13] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath. Optimal locally repairable and secure codes for distributed storage systems. preprint, arXiv:1210.6954, 2012.

[14] A. S. Rawat, A. Mazumdar, and S. Vishwanath. Cooperative local repair in distributed storage. *arXiv preprint arXiv:1409.3900*, 2014.

[15] N. B. Shah, K. Rashmi, and P. V. Kumar. Information-theoretically secure regenerating codes for distributed storage. In *Proc. IEEE GLOBECOM*. IEEE, 2011.

[16] N. B. Shah, K. Rashmi, and K. Ramchandran. Secure network coding for distributed secret sharing with low communication cost. In *Proc. Int. Symp. Inform. Theory*, pages 2404–2408. IEEE, 2013.

[17] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[18] I. Tamo and A. Barg. A family of optimal locally recoverable codes. *arXiv preprint arXiv:1311.3284*, 2013.

[19] R. Tandon and S. Mohajer. New bounds for distributed storage systems with secure repair. In *Proc. Allerton Conf. Commun., Contr., Computing*, 2014.