

Secret sharing schemes on graphs

László Csirmaz

Abstract

Given a graph G , a perfect secret sharing scheme based on G is a method to distribute a secret data among the vertices of G , the *participants*, so that a subset of participants can recover the secret if they contain an edge of G , otherwise they can obtain no information regarding the key. The average information rate is the ratio of the size of the secret and the average size of the share a participant must remember. The information rate of G is the supremum of the information rates realizable by perfect secret sharing schemes.

Based on the entropy-theoretical arguments due to Capocelli et al [3], and extending the results of M. van Dijk [6] we construct a graph G_n on n vertices with average information rate below $< 4/\log n$. We obtain this result by determining, up to a constant factor, the average information rate of the d -dimensional cube.

Key words. Secret sharing scheme, polymatroid, information theory.

1 Introduction

Secret sharing scheme is a method of distributing a secret data among a set of participants so that only qualified subsets are able to recover the data. If, in addition, unqualified subsets have no extra information, i.e. their joint shares is statistically independent of the secret, the scheme is called *perfect*. The goodness of a scheme is usually measured by how much information a participant must remember in the worst case, or in the average. Finding optimal perfect secret sharing schemes is important from both practical and theoretical point of view. Practically, the less information a participant must remember the more reliable the scheme is. Theoretically, the known upper and lower bounds are very far from each other, and closing the gap even in

some special cases is also an intriguing task. For a more complete description of the problem as well as a detailed list of references, see e.g. [1].

This paper is organized as follows. First we give the necessary definitions, and then state the theorems to be proved. Section 3 gives the proofs, and in Section 4 we conclude the paper. For undefined notions see [1] for secret sharing schemes, and [5] for those in information theory.

2 Definitions

In this section we give a rough definition of the notions we shall use later. First we define formally what a perfect secret sharing scheme is, then connect it to certain submodular function.

Let G be a graph, we denote the set of its vertices by V , and the number of the vertices by n . A subset A of V is *independent* or *stable*, if there is no edge between vertices in A . A *covering* of the graph G is a collection of subgraphs of G such that every edge of G is contained in one of the (not necessarily spanned) subgraphs in the collection. $K_{p,q}$ denotes the complete bipartite graph with disjoint classes of cardinality p and q . $K_{1,q}$ is called picturesquely as a *star*. For subsets of vertices we usually omit the \cup sign, and denote $A \cup B$ by AB . Also, if v is a vertex then Av denotes $A \cup \{v\}$. Finally, all logarithms in this paper are in base 2.

A *perfect secret sharing scheme* \mathcal{S} for a graph G is a collection of random variables ξ_v for $v \in V$ and ξ_s so that

- (i) ξ_s is the *secret*, and ξ_v is the *share* of v ;
- (ii) (edges can determine the secret) if there is an edge between v and $w \in V$ then ξ_v and ξ_w together determines the value of ξ_s ;
- (iii) (the scheme is perfect) if $A \subseteq V$ is independent, then ξ_s and the collection $\{\xi_v : v \in A\}$ are statistically independent.

We define the *size* of the random variable ξ as its *entropy*, or information content, denoted by $\mathbf{H}(\xi)$, cf. [5]. This is roughly the number of random bits necessary to determine the value of ξ . Thus the size of the secret is $\mathbf{H}(\xi_s)$, and the size of the share of $v \in V$ is $\mathbf{H}(\xi_v)$. The average size of the shares is $\sum_{v \in V} \mathbf{H}(\xi_v)/n$, and we are interested in the average information rate of \mathcal{S}

$$\tilde{\rho}_{\mathcal{S}} = \frac{n \cdot \mathbf{H}(\xi_s)}{\sum_{v \in V} \mathbf{H}(\xi_v)}.$$

For a given graph G its average information rate $\tilde{\rho} = \tilde{\rho}(G)$ is the supremum of $\tilde{\rho}_{\mathcal{S}}$ as \mathcal{S} runs over all possible perfect secret sharing schemes defined on G .

Claim 2.1 *For any graph G , $\tilde{\rho}(G)$ is at least $1/d$, where d is the average degree.*

Proof. We need to present a secret sharing scheme realizing the given rate. Let the secret be a single random unbiased bit s i.e. $\text{Prob}(s=0) = \text{Prob}(s=1) = 0.5$. For each edge (v, w) in G choose a random bit r independently of s and the previous choices, and tell the vertex v the bit r , and the vertex w the bit $r \oplus s$, their mod 2 sum. This is a perfect secret sharing scheme. Indeed, each edge can determine the secret, and an independent subset has a collection of independent random bits. The size of the secret is 1. For a vertex v , its share consists of independent random bits, one for each neighbor, thus the size of v 's share is the degree of v . Consequently the average size is the average degree, and we are done. ■

In the proof above G was covered by edges, i.e. by $K_{1,1}$ subgraphs. The construction can be generalized by using complete bipartite graphs instead of edges. For each $K_{p,q}$ in the covering choose a random bit r , give r to each member of one class, and $r \oplus s$ to members of the other class. Thus we have proved the following

Claim 2.2 *Suppose G is covered by complete bipartite graphs so that each vertex is covered λ times on the average. Then $\tilde{\rho}(G) \geq 1/\lambda$.* ■

Using stars instead of edges we can improve the bound in Claim 2.1. Orient all the edges of G arbitrarily. We have two stars at each vertex: one is formed by the incoming edges, the other by the outgoing edges. Suppose the secret consists of two bits, s_i and s_o , and distribute them as follows. For each vertex choose two random bits r_i and r_o , give them to the vertex, and give $r_i \oplus s_i$ to its incoming neighbors, and $r_o \oplus s_o$ to its outgoing neighbors. Each vertex gets one bit from each of its neighbor, plus two for its own, a total two more than its degree. This is for two bits of secret, which gives the average rate $2/(d+2)$ where d is the average degree. By a more sophisticated argument this still can be improved:

Theorem 2.3 (Stinson [10]) *For any graph G with average degree d , $\tilde{\rho}(G) \geq 2/(d+1)$.* ■

The following theorem, which we also quote without proof, shows that for dense graphs, i.e. graphs with average degree near to n , the previous bound is not the best possible.

Theorem 2.4 (Erdős and Pyber [7]) *There is a constant $c > 0$ so that for any graph G , $\tilde{\rho}(G) \geq c \log n/n$.* ■

In [7] it is proved that every graph can be covered by complete bipartite graphs so that each vertex is contained by at most $c \frac{n}{\log n}$ of the bipartite subgraphs. From here the theorem follows by Claim 2.2.

As J. Komlós observed [9], for random graphs the $c \frac{n}{\log n}$ bound is sharp, i.e. if in G every edge has probability $1/2$ then with high probability any cover of G with complete bipartite graphs has average cover number at least $n/4 \log n$. Therefore it seems a plausible conjecture that the average information rate of the random graph is also $c \frac{\log n}{n}$. Unfortunately the entropy method, presently the only available method for proving upper bounds, cannot give better estimate than $1/(1 + \alpha(G))$ where $\alpha(G)$ is the size of the maximal independent set, and this is $\log n$ for random graphs.

From the other side Stinson showed in [10] that the average information rate for any cycle of length ≥ 5 is exactly $2/3$. Capocelli et al [3] constructed access structures with information rate below $1/2 + \varepsilon$. For each $\varepsilon > 0$ and $d \geq 2$ van Dijk in [6] constructed a d -regular graph with information rate below $\varepsilon + 2/(d + 1)$. In this paper we present a graph with information rate below $1/\log n$. In fact we show that the d -dimensional cube, which is d -regular and has $n = 2^d$ vertices, has information rate between $2/\log n$ and $4/\log n$.

To prove upper bounds we use the entropy method introduced by Capocelli et al in [3], but also observed by others [8]. We recall the definitions and basic facts necessary to present the method. For a more detailed account on the method see e.g. [3] or [4].

Let \mathcal{S} be a perfect secret sharing scheme assigning random variables to the vertices of G . For subsets $A, B \subseteq V$ of the vertices let us define

$$f(A) \stackrel{\text{def}}{=} \frac{\mathbf{H}(\{\xi_v : v \in A\})}{\mathbf{H}(\xi_s)},$$

and

$$[[A, B]] \stackrel{\text{def}}{=} f(A) + f(B) - f(A \cup B) - f(A \cap B).$$

It is clear that $\tilde{\rho}_{\mathcal{S}} = \frac{1}{n} \sum_{v \in V} f(v)$. Using standard properties of the entropy function \mathbf{H} it is immediate that

- (i) $f(\emptyset) = 0$ and $f(A) \geq 0$;
- (ii) if $A \subseteq B \subseteq V$ then $f(B) \geq f(A)$;

(iii) $\llbracket A, B \rrbracket \geq 0$.

(ii) comes from the monotonicity of the entropy, and (iii) from the fact that the mutual conditional information is non-negative, cf. [5]. (iii) is referred to as submodularity, having the following immediate consequence:

$$f(A) + f(B) \geq f(A \cup B).$$

If ξ and η are independent random variables then $\mathbf{H}(\{\xi, \eta\}) = \mathbf{H}(\xi) + \mathbf{H}(\eta)$, and if ξ determines the value of η then $\mathbf{H}(\{\xi, \eta\}) = \mathbf{H}(\xi)$. Using these facts and the definition of the perfect secret sharing schemes, we have in addition,

Fact 2.5 (iv) *if $A \subseteq B \subseteq V$, A is an independent set of vertices, B is not, then $f(B) \geq 1 + f(A)$;*

(v) *if neither A nor B is independent, but $A \cap B$ is so, then $\llbracket A, B \rrbracket \geq 1$. ■*

Conditions (iv) and (v) are also sufficient in the following sense. If one assigns random variables to the secret and to the vertices of G , and the corresponding function f satisfies (iv) and (v), then this assignment constitutes a perfect secret sharing scheme.

After this introduction the entropy-method can be rephrased as follows. Since the function f arising from a perfect secret sharing scheme satisfies (i)–(v), if we prove that *any* function f satisfying (i)–(v) must take at least λ average value on the vertices, then $\tilde{\rho}(G) \leq 1/\lambda$. This is exactly what we shall do.

We omit the easy checking of the following two facts.

Fact 2.6 *For every pair A, B we have $\llbracket A, B \rrbracket = \llbracket B, A \rrbracket$. If $A \cap B \subseteq B' \subseteq B$, then*

$$\llbracket A, B \rrbracket = \llbracket A, B' \rrbracket + \llbracket AB', B \rrbracket.$$

In particular, $\llbracket A, B \rrbracket \geq \llbracket A, B' \rrbracket$. ■

Fact 2.7 *Suppose G_2 is a spanned subgraph of G_1 , and \mathcal{S}_1 is a perfect secret sharing scheme on G_1 . Keeping only the values corresponding to vertices in G_2 we get a perfect secret sharing scheme \mathcal{S}_2 on G_2 . Moreover, if A is a subset of vertices of the spanned subgraphs G_2 , then $f_{\mathcal{S}_2}(A) = f_{\mathcal{S}_1}(A)$. ■*

3 The result

The d -dimensional cube, denoted by C^d , has 2^d vertices which are labelled by 0-1 sequences of length d . Two vertices are connected by an edge if their

labels differ at exactly one position. The d -dimensional cube is d -regular, has two $d - 1$ -dimensional layers, each one is a $d - 1$ -dimensional cube, and there is a perfect matching between the layers. One can color the vertices of the cube in a chessboard-like fashion, showing that it is also a bipartite graph with equal classes of size 2^{d-1} . C_1 is an edge, C_2 a square, both has average information rate 1.

The definition of the lattice cube C_m^d is similar. The vertices are sequences of length d of integer numbers between 0 and $m - 1$. Two vertices are connected if the corresponding sequences differ in one position only and these numbers are consecutive ones. C_m^d consists of m layers of $d - 1$ -dimensional lattice cubes, bipartite with classes differing in size by at most one, and has average degree $2d(1 - 1/m)$. Thus by Stinson's result the average information rate for the d -dimensional lattice cube is at least $\frac{1}{d(1-1/m)}$.

Theorem 3.1 *The average information rate for the C_m^d lattice cube is less than $\frac{2}{d(1-1/m)}$.*

The next lemma is the key ingredient in our proof. Its application requires a large independent set, this explains while all constructions have small average degree.

Lemma 3.2 *Suppose A, B are disjoint subsets of the vertex set of G so that A is independent, B is not, and for each $a \in A$ there is a $b \in B$ which is connected to a only in A . Then*

$$\llbracket A, B \rrbracket \geq |A|.$$

Proof. Let $A = \{a_1, a_2, \dots, a_k\}$, and $b_j \in B$ which is connected to a_j only in A . Define $A_0 = \emptyset$, and for $j \leq k$ let $A_j = \{a_1, a_2, \dots, a_j\}$. Then by iterated application of Fact 2.6

$$\llbracket A, B \rrbracket = \llbracket A_k, B \rrbracket = \llbracket A_1, B \rrbracket + \llbracket A_2, A_1 B \rrbracket + \dots + \llbracket A_k, A_{k-1} B \rrbracket,$$

thus the lemma follows immediately from $\llbracket A_j, A_{j-1} B \rrbracket \geq 1$. By assumption, the subgraph $A_{j-1} b_j$ is independent, $A_j b_j$ and $A_{j-1} B$ are not, so by Fact 2.5 $\llbracket A_j b_j, A_{j-1} B \rrbracket \geq 1$, and then

$$\llbracket A_j, A_{j-1} B \rrbracket = \llbracket A_j, A_{j-1} b_j \rrbracket + \llbracket A_j b_j, A_{j-1} B \rrbracket \geq 0 + 1 = 1$$

as was claimed. ■

The next lemma is purely information-theoretical. It is a generalization of the identity $f(A) + f(B) = f(AB) + \llbracket A, B \rrbracket$ which holds whenever A and B are disjoint.

Lemma 3.3 *Suppose A_1, A_2, \dots, A_k are disjoint subsets of G . Then*

$$\sum_{1 \leq i \leq k} f(A_i) \geq f\left(\bigcup_{1 \leq i \leq k} A_i\right) + \sum_{1 \leq i < k} \llbracket A_i, A_{i+1} \rrbracket.$$

Proof. For $1 \leq i \leq k$ let $B_i = \bigcup_{j \leq i} A_j$, then $B_i \cup A_i A_{i+1} = B_{i+1}$. Using this, rearranging the terms in the sum

$$\llbracket B_2, A_2 A_3 \rrbracket + \llbracket B_3, A_3 A_4 \rrbracket + \dots + \llbracket B_{k-1}, A_{k-1} A_k \rrbracket + f(B_k) + \sum_{i < k} \llbracket A_i, A_{i+1} \rrbracket$$

one gets $\sum_{i \leq k} f(A_i)$. Since $\llbracket B_i, A_i A_{i+1} \rrbracket$ is always non-negative, the claim of the lemma follows. \blacksquare

Now we can prove Theorem 3.1. Suppose \mathcal{S} is a perfect secret sharing scheme on the d -dimensional lattice cube C_m^d , and $f = f_{\mathcal{S}}$ is the submodular function on the subsets of C_m^d defined from the entropy. We shall determine a sequence of constants λ_d so that the following inequality holds:

$$\sum_{v \in C_m^d} f(v) \geq f(C_m^d) + \lambda_d \cdot |C_m^d|, \quad (1)$$

where λ_d will depend on m but not on the scheme \mathcal{S} . The idea of considering this inequality comes from van Dijk [6].

For the case $d = 1$ we have a path of length m , let the vertices of the graph be $V = \{v_1, \dots, v_m\}$ in this order. Let $A_1 = v_1$, $A_2 = v_2 v_3$, $A_3 = v_3$, $A_4 = v_4 v_5$, etc. alternately grouping one and two vertices. By the subadditivity

$$\sum_{v \in C_m^1} \geq f(A_1) + f(A_2) + \dots \geq f(C_m^1) + \sum \llbracket A_i, A_{i+1} \rrbracket$$

applying Lemma 3.3. By Lemma 3.2 a member on the right hand side sum is ≥ 1 if at least one of A_i and A_{i+1} has two connected elements. If $m > 2$ then there are at least $\frac{2}{3}(m-1)$ such a pair, thus

$$\sum_{v \in C_m^1} \geq f(C_m^1) + \frac{2}{3}(m-1) > f(C_m^1) + \frac{m-1}{2m} |C_m^1|.$$

This means that for $m > 2$ (1) holds with $\lambda_1 = \frac{m-1}{2m}$; for $m = 2$ we choose $\lambda_1 = 0$.

Now suppose we know that (1) holds for d with some λ_d , and we would like to determine λ_{d+1} . The $d+1$ -dimensional cube C_m^{d+1} consists of m layers, each of them is a C_m^d cube. As Fact 2.7 shows, restricting the scheme \mathcal{S} to any of the subcubes induces a subscheme with the same f function. Denoting the j -th layer by A_j , thus we can apply (1) to A_j which yields

$$\sum_{v \in A_j} f(v) \geq f(A_j) + \lambda_d \cdot |C_m^d|,$$

since A_j is a d -dimensional lattice cube. Adding up these inequalities and using Lemma 3.3 to the disjoint subsets A_1, \dots, A_m we get

$$\sum_{v \in C_m^{d+1}} f(v) \geq f(C_m^{d+1}) + \sum_{1 \leq j < m} \llbracket A_j, A_{j+1} \rrbracket + m\lambda_d |C_m^d|.$$

Next we estimate $\llbracket A_j, A_{j+1} \rrbracket$. In C_m^{d+1} there is a perfect matching between A_j and A_{j+1} , and neither A_j nor A_{j+1} is independent. Suppose $B \subseteq A_j$ is a maximal independent set, then Fact 2.6 and Lemma 3.2 gives

$$\llbracket A_j, A_{j+1} \rrbracket \geq \llbracket B, A_{j+1} \rrbracket \geq |B| \geq \frac{1}{2} |C_m^d|,$$

as the maximal independent set in C_m^d has size $\lfloor (|C_m^d| + 1)/2 \rfloor$. Thus

$$\sum_{v \in C_m^{d+1}} f(v) \geq f(C_m^{d+1}) + \frac{m-1}{2} |C_m^d| + m\lambda_d |C_m^d|.$$

Since $m|C_m^d| = |C_m^{d+1}|$, with $\lambda_{d+1} = \lambda_d + \frac{m-1}{2m}$ inequality (1) will also hold for $d+1$.

As $f(C_m^d) > 0$, from (1) we can conclude immediately that the average information rate of C_m^d is less than $1/\lambda_d$. We have seen that for $m > 2$ $\lambda_d = \frac{m-1}{2m}$ is a good choice and this proves Theorem 3.1 in this case.

For the case $m = 2$ this gives only $\lambda_d = (d-1)/4$. The missing $1/4$ will come from the first member $f(C^d)$ in (1). The d -dimensional cube has two $d-1$ -dimensional layers A_1 and A_2 , and we have seen that

$$\llbracket A_1, A_2 \rrbracket \geq \frac{1}{2} |C^{d-1}| = \frac{1}{4} |C^d|.$$

Since $f(C^d) = f(A_1 A_2) \geq \llbracket A_1, A_2 \rrbracket$ this proves Theorem 3.1 for $m = 2$ too.

■

In terms of number of vertices, the best (i.e. smallest) value is $4/\log n$ given by $m = 2$, compared to $4.755/\log n$ and $5.33/\log n$ when $m = 3$ or $m = 4$.

In the reasoning above we have used the existence half of the edges only between the consecutive layers of C_m^{d+1} . This means that we can simply remove these edges without affecting the upper bound but almost halving the average degree. For this truncated graph the upper and lower bounds are practically equal.

4 Conclusion

We have presented a graph with average information rate below $4/\log n$. The best general construction of [7] shows that we cannot go below $\frac{\log n}{n}$.

Problem 4.1 *Do there exist graphs G_n on n vertices with average information rate $(\omega(n) \log n)^{-1}$ such that $\omega(n)$ tends to infinity?*

Our graph was the d -dimensional cube. For $d = 1$ and $d = 2$ we know the exact values: both of them has average information rate 1. The next case is the 3-dimensional cube. This is 3-regular, thus Stinson's theorem gives the lower bound $1/2$. Our theorem gives the upper bound $4/3$ which is worthless since 1 is also an upper bound. By a result of [3] the two middle points of a path consisting three edges have average information rate at most $2/3$. Observing that each edge of the cube can be the middle segment of a spanned path of length three, we get that the average information rate of the cube cannot exceed $2/3$.

Problem 4.2 *Determine the average information rate of the 3-dimensional cube. It is between $1/2$ and $2/3$.*

We remark that $2/3$ is the limit of the entropy method, since there exists a function f on the subset of vertices satisfying (i)-(v) of Section 2 with $f(v) = 3/2$ for all vertices.

In general the lower and upper bounds differ by a factor of 2, for C_d the average information rate is between $c_d/\log n$ where $2 \leq c_d \leq 4$. It is not clear that c_d should converge, however we conjecture that it does.

Problem 4.3 *Determine the constant c in the asymptotic information rate $c/\log n$ of the d -dimensional cube C^d .*

We have seen that there is a graph G_n on n vertices with information rate $\Omega(1/\log n)$. We know that the random graph on n vertices contains as spanned subgraph every possible graph on $\log n$ vertices, thus $G_{\log n}$ too. It means that the information rate of some vertex in the random graph is $\Omega(1/\log \log n)$. Unfortunately this gives very poor estimate on the average information rate.

Problem 4.4 *Prove that the average information rate of the random graph is less than $c/\log n$ for some constant c .*

The best upper bound on the average as well as on the worst case information rate is above $1/(\alpha(G) + 1)$ as the following reasoning shows. For each subset A of the vertices define $f(A)$ as a function on the number of vertices in A as follows. If $|A| = 1$ then $f(A) = 1 + \alpha(G)$; if $v \notin A$ and $1 \leq |A| \leq \alpha(G)$ then $f(Av) = f(A) + 1 + \alpha(G) - |A|$, and if $\alpha(G) < |A|$ then $f(Av) = f(A)$. It is a routine to check that this f satisfies conditions (i)-(v) of Section 2, therefore it will met any bound using the entropy inequalities only. For the random graph $\alpha(G) = \log n$, thus the entropy method cannot give better result than what Problem 4.4 requires.

Problem 4.5 *Determine the average information rate of the random graph, or at least prove that it is $\omega(n)/\log n$ where $\omega(n)$ tends to zero.*

References

- [1] C. Blundo, A. De Santis, D. R. Stinson, U. Vaccaro, Graph Decomposition and Secret Sharing Schemes *Journal of Cryptology*, Vol 8(1995) pp. 39–64.
- [2] E. F. Brickell and D. R. Stinson, Some improved bounds on the information rate of perfect secret sharing schemes *Journal of Cryptology*, Vol 5(1992) pp. 153–166.
- [3] R. M. Capocelli, A. De Santis, L. Gargano, U. Vaccaro, On the Size of Shares for Secret Sharing Schemes, *Journal of Cryptology*, Vol 6(1993) pp. 157-168.
- [4] L. Csirmaz, The size of a share must be large, *Journal of Cryptology*, to appear
- [5] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.
- [6] M. van Dijk, On the Information Rate of Perfect Secret Sharing Schemes, preprint, 1994

- [7] P. Erdős, L. Pyber, Covering a graph by complete bipartite graphs, preprint, 1995
- [8] K. Kurosawa, K. Okada, K. Sakano, W. Ogata, S. Tsujii, Nonperfect Secret Sharing Schemes and Matroids, *Proceedings of Eurocrypt'93*.
- [9] J. Komlós, personal communication.
- [10] D. R. Stinson, Decomposition construction for secret sharing schemes, *IEEE Trans. Inform. Theory* Vol 40(1994) pp. 118-125.