On the Information Rate of Secret Sharing Schemes*

Extended Abstract

C. Blundo, A. De Santis, L. Gargano, U. Vaccaro

Dipartimento di Informatica ed Applicazioni, Università di Salerno
84081 Baronissi (SA), Italy

Abstract. We derive new limitations on the information rate and the average information rate of secret sharing schemes for access structure represented by graphs. We give the first proof of the existence of access structures with optimal information rate and optimal average information rate less that $1/2 + \epsilon$, where ϵ is an arbitrary positive constant. We also provide several general lower bounds on information rate and average information rate of graphs. In particular, we show that any graph with n vertices admits a secret sharing scheme with information rate $\Omega((\log n)/n)$.

1 Introduction

A secret sharing scheme is a technique to distribute a secret S among a set of participants P in such a way that only qualified subsets of P can reconstruct the value of S whereas any other subset of P, non-qualified to know S, cannot determine anything about the value of the secret. We briefly recall the results on secret sharing schemes that are more closely related to the topics of this paper.

Shamir [19] and Blackley [2] were the first to consider the problem of secret sharing and gave secret sharing schemes where each subset A of P of size $|A| \geq k$ can reconstruct the secret, and any subset A of participants of size $|A| < k$ have absolutely no information on the secret. These schemes are known as (n, k) *threshold schemes*; the value k is the threshold of the scheme and n is the size of P.

Ito, Saito and Nishizeki [15] considered a more general framework and showed how to realize a secret sharing scheme for any access structure. An access structure is a family of all subsets of P which are qualified to recover the secret. Their technique requires that the size of set where the shares are taken be very large compared to the size of the set where the secret is chosen. Benaloh and Leichter [1] proposed a technique to realize a secret sharing scheme for any access structure more efficient than Ito, Saito and Nishizeki's methodology. It should be pointed out that threshold schemes are insufficient to realize a secret sharing

* Partially supported by Italian Ministry of University and Research (M.U.R.S.T.) and by National Council for Research (C.N.R.) under grant 91.02326.CT12.

E.F. Brickell (Ed.): Advances in Cryptology - CRYPTO '92, LNCS 740, pp. 148-167, 1993.

scheme for general access structures $\mathcal{A}$ [1]. Moreover, Benaloh and Leichter also showed that there exist access structures for which any secret sharing scheme must give to some participant a share which is from a domain strictly larger than that of the secret.

Brickell and Davenport [5] analyzed ideal secret sharing schemes in terms of matroids. An ideal secret sharing scheme is a scheme for which the the shares are taken has the same size of the set where the secret is chosen. In particular, they proved that an ideal secret sharing scheme exists for a graph G, if and only if G is a complete multipartite graph. Equivalently, if we define the information rate as the ratio between the size of the secret and that of the biggest share given to any participant, Brickell and Davenport's result can be stated saying that a graph has information rate 1 if and only if it is a complete multipartite graph. Brickell and Stinson [6] gave several upper and lower bounds on the information rate of access structures based on graphs.

Capocelli, De Santis, Gargano, and Vaccaro [7] gave the first example of access structures with information rate bounded away from 1.

Blundo, De Santis, Stinson, and Vaccaro [4] analyzed the information rate and the average information rate of secret sharing schemes based on graphs. The average information rate is the ratio between the secret size and the arithmetic mean of the size of the shares for such schemes. They proved the existence of a gap in the values of information rates of graphs, more precisely they proved that if a graph G with n vertices is not a complete multipartite graph then any secret sharing scheme for it has information rate not greater than $2/3$ and average information rate not greater than $n/(n+1)$. These upper bounds arise by applying entropy argument due to Capocelli, De Santis, Gargano, and Vaccaro [7].

The recent survey by Stinson [21] contains an unified description of recent results in the area of secret sharing schemes. For different approaches to the study of secret sharing schemes, for schemes with "extended capabilities" as dis-enrollment, fault-tolerance, and pre-positioning and for a complete bibliography we recommend the survey article by Simmons [20].

In this paper we derive new limitations on the information rate and the average information rate for access structures represented by graphs. In the first part we prove new upper bounds on the information rate and the average information rate. These bounds are obtained by using the entropy approach by [7] and are the best possible for the considered structures since we exhibit secret sharing schemes that meet the bounds. In particular, we give the first proof of the existence of access structures with information rate and average information rate strictly less that $2/3$. This solves a problem of [4]. In the second part we consider the problem of finding good lower bounds on the information rate and the average information rate and we give several general lower bounds that improve on previously known results.

2 Preliminaries

In this section we review the basic concepts of Information Theory we shall use. For a complete treatment of the subject the reader is advised to consult [8] and [11]. We shall also recall some basic terminology from graph theory.

Given a probability distribution $\{p(x)\}_{x \epsilon X}$ on a set X, we define the *entropy* of X, $H(X)$, as

$$H(X) = -\sum_{x \epsilon X} p(x) \log p(x)^2.$$

The entropy $H(X)$ is a measure of the average uncertainty one has about which element of the set X has been chosen when the choices of the elements from X are made according to the probability distribution $\{p(x)\}_{x \epsilon X}$. The entropy enjoys the following property

$$0 \leq H(X) \leq \log |X|, \tag{1}$$

where $H(X) = 0$ if and only if there exists $x_0 \in X$ such that $p(x_0) = 1$; $H(X) = \log |X|$ if and only if $p(x) = 1/|X|$, for all $x \in X$.

Given two sets X and Y and a joint probability distribution $\{p(x,y)\}_{x \epsilon X, y \epsilon Y}$ on their Cartesian product, the *conditional entropy* $H(X|Y)$, also called the equivocation of X given Y, is defined as

$$H(X|Y) = -\sum_{y \epsilon Y} \sum_{x \epsilon X} p(y)p(x|y) \log p(x|y).$$

The conditional entropy can be written as

$$H(X|Y) = \sum_{y \epsilon Y} p(y) H(X|Y=y)$$

where $H(X|Y=y) = -\sum_{x \epsilon X} p(x|y) \log p(x|y)$. From the definition of conditional entropy it is easy to see that

$$H(X|Y) \geq 0. \tag{2}$$

If we have $n+1$ sets $X_1, \ldots, X_n, Y$, the entropy of $X_1 \ldots X_n$ given Y can be expressed as

$$H(X_1 \ldots X_n|Y) = H(X_1|Y) + H(X_2|X_1 Y) + \cdots + H(X_n|X_1 \ldots X_{n-1} Y) \tag{3}$$

The *mutual information* between X and Y is defined by

$$I(X;Y) = H(X) - H(X|Y) \tag{4}$$

and enjoys the following properties:

$$I(X;Y) = I(Y;X), \tag{5}$$

[2] All logarithms in this paper are of base 2

and

$$I(X;Y) \geq 0,$$

from which one gets

$$H(X) \geq H(X|Y). \tag{6}$$

Given $n+2$ sets $X, Y, Z_1, \ldots, Z_n$ and a joint probability distribution on their Cartesian product, the *conditional mutual information* between X and Y given $Z_1, \ldots, Z_n$ can be written as

$$I(XY|Z_1, \ldots, Z_n) = H(X|Z_1, \ldots, Z_n) - H(X|Z_1, \ldots, Z_nY). \tag{7}$$

Since the conditional mutual information is always non negative we get

$$H(X|Z_1, \ldots, Z_n) \geq H(X|Z_1, \ldots, Z_nY). \tag{8}$$

We now present some basic terminology from graph theory. A graph, $G = (V(G), E(G))$ consists of a finite non empty set of vertices $V(G)$ and a set of edges $E(G) \subseteq V(G) \times V(G)$. Graphs do not have loops or multiple edges. We consider only undirected graphs. In an undirected graph the pair of vertices representing any edge is unordered. Thus, the pairs (X, Y) and (Y, X) represent the same edge. To avoid overburdening the notation we often describe a graph G by the list of all edges $E(G)$. We will use reciprocally (X, Y) and XY to denote the edge joining the vertices X and Y. G is *connected* if any two vertices are joined by a path. The *complete graph* K_n is the graph on n vertices in which any two vertices are joined by an edge. The *complete multipartite graph* $K_{n_1,n_2,\ldots,n_t}$ is a graph on $\sum_{i=1}^{t} n_i$ vertices, in which the vertex set is partitioned into subsets of size n_i $(1 \leq i \leq t)$ called *parts*, such that vw is an edge if and only if v and w are in different parts.

Suppose G is a graph and $G_1, \ldots, G_t$ are subgraphs of G, such that each edge of G occurs in at least one of the G_i's. We say that $\Pi = \{G_1, \ldots, G_t\}$ is a covering of G and if each G_i, $i = 1, \ldots, t$ is a complete multipartite graph then we say that Π is a *complete multipartite covering* (CMC) of G.

3 Secret Sharing Schemes

A secret sharing scheme permits a secret to be shared among n participants in such a way that only qualified subsets of them can recover the secret, but any non-qualified subset has absolutely no information on the secret. An access structure $\mathcal{A}$ is the set of all subsets of P that can recover the secret.

Definition 1. Let P be a set of participants, a monotone access structure $\mathcal{A}$ on P is a subset $\mathcal{A} \subseteq 2^P$, such that

$$A \in \mathcal{A}, A \subseteq A' \subseteq P \Rightarrow A' \in \mathcal{A}.$$

Definition 2. Let P a set of participants and $A \subseteq 2^P$. The closure of A, *cl*(A), is the set

$$cl(A) = \{C | B \in A \text{ and } B \subseteq C \subseteq P\}.$$

For a monotone access structure $\mathcal{A}$ we have $\mathcal{A} = cl(\mathcal{A})$.

A secret sharing scheme for secrets $s \in S$ and a probability distribution $\{p(s)\}_{s \in S}$ naturally induce a probability distribution on the joint space defined by the shares given to participants. This specifies the probability that participants receive given shares.

In terms of the probability distribution on the secret and on the shares given to participants, we say that a secret sharing scheme is a *perfect* secret sharing scheme, or simply a secret sharing scheme, for the monotone access structure $\mathcal{A} \subseteq 2^P$ if

1. *Any subset $A \subseteq P$ of participants not enabled to recover the secret have no information on the secret value:*[3]
   If $A \notin \mathcal{A}$ then for all $s \in S$ and for all $a \in A$ it holds $p(s|a) = p(s)$.
2. *Any subset $A \subseteq P$ of participants enabled to recover the secret can compute the secret:*
   If $A \in \mathcal{A}$ then for all $a \in A$ a unique secret $s \in S$ exists such that $p(s|a) = 1$.

Notice that the property 1. means that the probability that the secret is equal to s given that the shares held by $A \notin \mathcal{A}$ are a, is the same of the *a priori* probability that the secret is s. Therefore, no amount of knowledge of shares of participants not enabled to reconstruct the secret enables a Bayesian opponent to modify an *a priori* guess regarding which the secret is. Property 2. means that the value of the shares held by $A \in \mathcal{A}$ univocally determines the secret $s \in S$.

Let P be a set of participants, and $\mathcal{A}$ be a monotone access structure on P. Following the approach of [13], [14], and [7] we can restate above conditions 1. and 2. using the information measures introduced in the previous section. Therefore, we say that a secret sharing scheme is a sharing of the secret S among participants in P such that

1′. *Any qualified subset can reconstruct the secret.*
   Formally, for all $A \in \mathcal{A}$, it holds $H(S|A) = 0$.
2′. *Any non-qualified subset has absolutely no information on the secret.*
   Formally, for all $A \notin \mathcal{A}$, it holds $H(S|A) = H(S)$.

Notice that $H(S|A) = 0$ means that each set of values of the shares in A corresponds to a unique value of the secret. In fact, by definition, $H(S|A) = 0$ is equivalent to the fact that for all $a \in A$ with $p(a) \neq 0$ exists $s \in S$ such that $p(s|a) = 1$. Moreover, $H(S|A) = H(S)$ is equivalent to state that S and A are statistically independent, i.e., for all $a \in A$ for all $s \in S$, $p(s|a) = p(s)$ and therefore the knowledge of a gives no information about the secret. Notice that the condition $H(S|A) = H(S)$ is equivalent to say that for all $a \in A$ it holds $H(S|A = a) = H(S)$.

[3] To maintain notation simpler, we denote with the same symbol (sets of) participant(s) and the set(s) from which their shares are taken.

3.1 The Size of the Shares

One of the basic problems in the field of secret sharing schemes is to derive bounds on the amount of information that must be kept secret. This is important from the practical point of view since the security of any system degrades as the amount of secret information increases.

Let P be a set of n participants and $\mathcal{A} \subseteq 2^P$ be an access structure on P. We denote by $X \in P$ either the participant X or the random variable defined by the value of his share. Different measures of the amount of secret information that must be distributed in a secret sharing scheme are possible. If we are interested in limiting the maximum size of shares for each participant (i.e., the maximum quantity of secret information that must be given to any participant), then a worst-case measure of the maximum of $H(X)$ over all $X \in P$ naturally arises. To analyze such cases we use the *information rate* of $\mathcal{A}$ defined as

$$\rho(\mathcal{A}, \mathcal{P}_S) = \frac{H(S)}{\max_{X \epsilon P} H(X)},$$

for a given secret sharing scheme and non-trivial probability distribution $\mathcal{P}_S$ on the secret. This measure was introduced by Brickell and Stinson [6] when the probability distributions over the secret and the shares are uniform. In such a case the definition becomes $\rho(\mathcal{A}) = \log|S|/\max_{X \epsilon P} \log|X|$. The optimal information rate is then defined as:

$$\rho^*(\mathcal{A}) = \sup_{\mathcal{T}, \mathcal{Q}} \frac{H(S)}{\max_{X \epsilon P} H(X)},$$

where $\mathcal{T}$ is the space of all secret sharing schemes for the access structure $\mathcal{A}$ and $\mathcal{Q}$ is the space of all non-trivial probability distributions $\mathcal{P}_S$.

In many cases it is preferable to limit the sum of the size of shares given to all participants. In such a case the arithmetic mean of the $H(X)$, $X \in P$, is a more appropriate measure. We define the *average information rate* as follows

$$\widetilde{\rho}(\mathcal{A}, \mathcal{P}_S) = \frac{H(S)}{\sum_{X \epsilon P} H(X)/|P|},$$

for a given secret sharing scheme and non-trivial probability distribution $\mathcal{P}_S$ on the secret. This measure was introduced in [3], [16], and [17] when an uniform probability distribution on the set of secrets is assumed. Blundo, De Santis, Stinson, and Vaccaro [4] analyzed secret sharing schemes by means of this measure, when the probability distributions over the secret and the shares are uniform. If the secret and the shares are chosen under a uniform probability distribution, considering previous measure is equivalent to consider the "average size" of the shares assigned to each participant to realize a secret sharing scheme. The optimal average information rate is then defined as:

$$\widetilde{\rho}^*(\mathcal{A}) = \sup_{\mathcal{T}, \mathcal{Q}} \frac{H(S)}{\sum_{X \epsilon P} H(X)/|P|}.$$

It is clear that, for the same secret sharing scheme and non-trivial probability distribution $\mathcal{P}_S$ on the secret, the information rate is no greater than the average information rate, that is $\widetilde{\rho} \geq \rho$ and $\widetilde{\rho} = \rho$ if and only if all $H(X)$, $X \in P$, have the same value. As done in [4] we denote, for a graph G, the optimal information rate with $\rho^*(G)$ and the average information rate with $\widetilde{\rho}^*(G)$.

3.2 Auxiliary Results

In this section we recall some auxiliary results. We will improve some of them in the next sections and we will use others in our constructions.

Brickell and Stinson [6] proved the following lower bound on the information rate for any graph of maximum degree d.

Theorem 3. *Let G be a graph with maximum degree d, then*

$$\rho^*(G) \geq \frac{1}{\lceil d/2 \rceil + 1}.$$

In Section 4 we will show how to improve on it for odd d. Blundo, De Santis, Stinson, and Vaccaro [4] proved the following results for acyclic graphs

Lemma 4. *Let G be a tree, then a secret sharing scheme for G exists with information rate equal to $1/2$. Thus $\rho^*(G) \geq 1/2$.*

In Section 4 we will show how to improve this bound for any tree.

The following result, proved in [4] will be used to obtain good secret sharing schemes for graphs with maximum degree 3.

Theorem 5. *Let P_n be a path of length n, $n \geq 3$. A secret sharing scheme for P_n exists with optimal information rate $2/3$.*

The following lemmas have been proved by Capocelli, De Santis, Gargano, and Vaccaro [7]; we will use them to find new upper bounds on the information rate of access structures. Since their proofs are simple, we report them for reader's convenience.

Lemma 6. *Let $\mathcal{A}$ be an access structures on a set P of participants and $X, Y \subset P$. Let $Y \notin \mathcal{A}$ and $X \cup Y \in \mathcal{A}$. Then $H(X|Y) = H(S) + H(X|YS)$.*

Proof. The conditional mutual information $I(X;S|Y)$ can be written either as $H(X|Y) - H(X|YS)$ or as $H(S|Y) - H(S|XY)$. Hence, $H(X|Y) = H(X|YS) + H(S|Y) - H(S|XY)$. Because of $H(S|XY) = 0$ for $X \cup Y \in \mathcal{A}$ and $H(S|Y) = H(S)$ for $Y \notin \mathcal{A}$, we have $H(X|Y) = H(S) + H(X|YS)$. □

Lemma 7. *Let $\mathcal{A}$ an access structures on a set P of participants and $X, Y \subset P$. If $X \cup Y \notin \mathcal{A}$ then $H(Y|X) = H(Y|XS)$.*

Proof. The conditional mutual information $I(Y, S|X)$ X can be written either as $H(Y|X) - H(Y|XS)$ or as $H(S|X) - H(S|XY)$. Hence, $H(Y|X) = H(Y|XS) + H(S|X) - H(S|XY)$. Because of $H(S|XY) = H(S|X) = H(S)$, for $X \cup Y \notin \mathcal{A}$, we have $H(Y|X) = H(Y|XS)$. □

Finally, we briefly recall a technique introduced in [4] to obtain lower bounds on the information rate of a graph G.

Suppose G is a graph and $G_1, \ldots, G_n$ are subgraphs of G, such that each edge of G occurs in at least one of the G_i's. Suppose also that each G_i is a complete multipartite graph. Then we say that $\Pi = \{G_1, \ldots, G_t\}$ is a *complete multipartite covering* (or CMC) of G. Let $\Pi_j = \{G_{j1}, \ldots, G_{jn_j}\}$, $j = 1, \ldots L$, comprise a complete enumeration of the minimal CMCs of G. For every vertex v and for $j = 1, \ldots L$ define $R_{jv} = |\{i : v \in G_{ji}\}|$ and consider the following optimization problem $\mathcal{O}(G)$:

$$\boxed{\begin{array}{ll} \text{Minimize } T \text{ subject to:} & \\ & a_j \geq 0, \;\; 1 \leq j \leq L \\ & \sum_{j=1}^{L} a_j = 1 \\ & T \geq \sum_{j=1}^{L} a_j R_{jv}, \;\; v \in V(G) \end{array}}$$

In citeBlDeStVa it is proved that if T^* is the optimal solution to $\mathcal{O}(G)$ then $\rho^*(G) \geq 1/T^*$.

4 Upper Bounds on the Information Rate and Average Information Rate

In this section we will exhibit an access structure having information rate less than 2/3. This solves an open problem in [4]. The result is obtained using the entropy approach of [7].

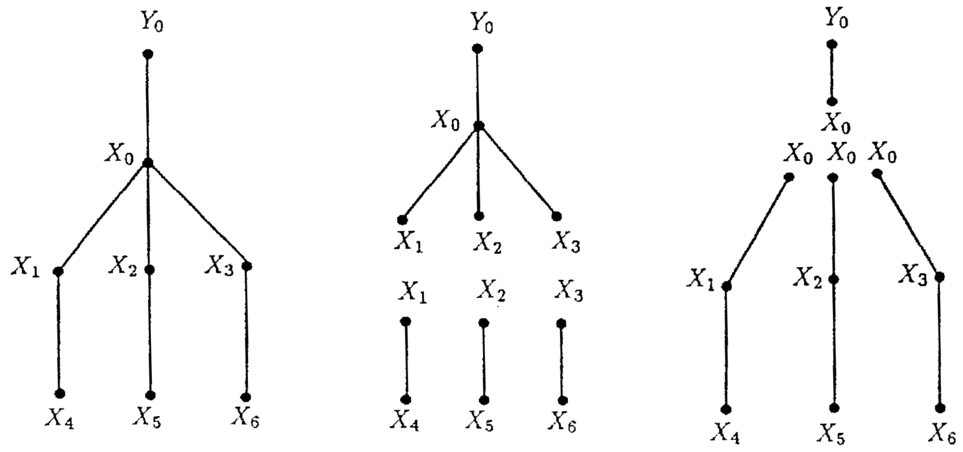
Consider the graph $\mathcal{AS}_k = (V(\mathcal{AS}_k), E(\mathcal{AS}_k))$, $k \geq 1$, where

$$V(\mathcal{AS}_k) = \{Y_0, X_0, X_1, \ldots, X_k, X_{k+1}, \ldots, X_{2k}\}$$

and

$$E(\mathcal{AS}_k) = \{(Y_0, X_0), (X_0, X_1), \ldots, (X_0, X_k), (X_1, X_{k+1}), \ldots, (X_k, X_{2k})\}.$$

As an example, the graph $\mathcal{AS}_k$ for $k = 3$ is depicted in Figure $1(a)$.



(a) (b) (c)

Figure 1

Theorem 8. *The optimal information rate of the graph AS_k, $k \geq 1$, is*

$$\rho^*(AS_k) = \frac{1}{2} + \frac{1}{4k+2},$$

and the optimal average information rate is

$$\widetilde{\rho}^*(AS_k) = \frac{2}{3} + \frac{2}{9k+6}.$$

Proof: Consider the conditional entropy $H(X_1 \dots X_k | Y_0)$. We have

$$\begin{aligned} H(X_1 \dots X_k|Y_0) &= H(X_1|Y_0) + H(X_2|X_1Y_0) + \cdots + H(X_k|X_1 \dots X_{k-1}Y_0) \quad \text{(from (3))} \\ &\geq H(X_1|Y_0X_{k+1}) + H(X_2|X_1Y_0X_{k+2}) + \\ &\quad H(X_3|X_1X_2Y_0X_{k+3}) + \cdots + H(X_k|X_1 \dots X_{k-1}Y_0X_{2k}) \quad \text{(from (8))} \\ &\geq kH(S) \quad \text{(from Lemma 6 and (2)).} \end{aligned}$$

On the other hand, we have also

$$\begin{aligned} H(X_1 \dots X_k|Y_0) &= H(X_1 \dots X_k|Y_0S) \quad \text{(from Lemma 7)} \\ &\leq H(X_0X_1 \dots X_k|Y_0S) \quad \text{(from (3) and (2))} \\ &\leq H(X_0|Y_0S) + H(X_1|X_0S) + \cdots + H(X_k|X_0S) \quad \text{(from (3) and (8))} \\ &= H(X_0|Y_0) - H(S) + \cdots + H(X_k|X_0) - H(S) \quad \text{(from Lemma 6)} \\ &\leq H(X_0) + \cdots + H(X_k) - (k+1)H(S) \quad \text{(from (6)).} \end{aligned}$$

Therefore, we get

$$H(X_0) + H(X_1) + \ldots + H(X_k) \geq (2k+1)H(S). \tag{9}$$

From (9) it follows that there exists $i \in \{0, 1, \ldots, k\}$ such that

$$H(X_i) \geq \frac{2k+1}{k+1} H(S).$$

Therefore, the optimal information rate of $\mathcal{AS}_k$ $\rho^*(\mathcal{AS}_k)$ is upper bounded by

$$\rho^*(\mathcal{AS}_k) = \frac{H(S)}{\max H(X)} \leq \frac{k+1}{2k+1} = \frac{1}{2} + \frac{1}{4k+2}.$$

From (9) and from Lemma 6 it follows that

$$H(Y_0) + \sum_{i=0}^{2k} H(X_i) \geq (3k+2) H(S).$$

Therefore, the optimal average information rate of $\mathcal{AS}_k$ is upper bounded by

$$\frac{2k+2}{3k+2} = \frac{2}{3} + \frac{2}{9k+6}.$$

Actually, $1/2 + 1/(4k+2)$ is the true value of the optimal information rate. This value can be attained by using the CMC technique presented in [4] as solution of the following linear programming problem.
Consider the following two minimal complete multipartite coverings of $\mathcal{AS}_k$

$$\Pi_1 = \Big\{\{Y_0X_0, X_0X_1, \cdots, X_0X_k\}, \{X_1X_{k+1}, \cdots, X_kX_{2k}\}\Big\}$$
$$\Pi_2 = \Big\{\{Y_0X_0\}, \{X_0X_1, X_1X_{k+1}\}, \cdots, \{X_0X_k, X_kX_{2k}\}\Big\}.$$

An example of these two covering of $\mathcal{AS}_k$ are depicted in Figure 1(b) and 1(c) for $k = 3$. The matrix of entries R_{jv} is

$$\begin{pmatrix} 1 & 1 & \overbrace{2\cdots2}^{k}\ \overbrace{1\cdots1}^{k} \\ 1 & k+1 & \underbrace{1\cdots\cdots\cdots1}_{2k} \end{pmatrix}.$$

Hence the linear programming problem to be solved is the following:

Minimize T subject to

$$\begin{aligned} a_j &\geq 0,\ j = 1, 2 \\ a_1 + a_2 &= 1 \\ T &\geq a_1 + (k+1) a_2 \\ T &\geq 2a_1 + a_2 \end{aligned}$$

The optimal solution is

$$(a_1, a_2, T) = \left(\frac{k}{k+1}, \frac{1}{k+1}, \frac{2k+1}{k+1}\right).$$

Hence, $\rho_C^*(\mathcal{AS}_k) = (2k+1)/(k+1)$, and this rate can be attained by taking k copies of Π_1, and one copy of Π_2. Thus, the optimal information rate of $\mathcal{AS}_k$ is $1/2 + 1/(4k+2)$. The optimal average information rate equal to $2/3 + 2/(9k+6)$ can be attained by either Π_1 or Π_2. □

Suppose that $p(s) = 1/|S|$, for any $s \in S$. Above result and inequality (1) imply that any perfect secret sharing scheme for $\mathcal{AS}_k$ must give to at least a participant a share of size greater than $2 - 1/(k+1)$ times the size of the secret.

Theorem 8 is a generalization of Theorem 4.1 of [7]. In fact if we choose $k = 1$ the access structure $\mathcal{AS}_k$ is the closure of the edge-set of P_3, the path on four vertices.

In Appendix A are depicted all graphs on six vertices that have $\mathcal{AS}_2$ as induced subgraph and, therefore, have optimal information rate less than 3/5. It turns out that the optimal information rate for all those graphs is equal to 3/5, and all but one have also an optimal average information rate equal to 3/4.

Using the previous theorem we can show the existence of access structures having *average information* rate less than 2/3, which represented the best upper bound known so far [7]. Consider the graph $\mathcal{M}_k$, where $V(\mathcal{M}_k) = \{X_1, X_2, \ldots, X_{2k+3}, X_{2k+4}\}$ and

$$E(\mathcal{M}_k) = \{X_1X_2\} \bigcup \{X_2X_i, X_iX_{k+i}, X_{k+i}X_{2k+3} | 3 \leq i \leq k+2\} \bigcup \{X_{2k+3}X_{2k+4}\}.$$

The graph $\mathcal{M}_3$ is depicted in Figure 2. The following theorem holds.

Theorem 9. *The optimal average information rate for $\mathcal{M}_k$, $k \geq 1$, is*

$$\widetilde{\rho}^*(\mathcal{M}_k) = \frac{k+2}{2k+2}.$$

Proof : From Lemma 6 we get $H(X_1) \geq H(S)$ and $H(X_{2k+4}) \geq H(S)$, whereas from Theorem 8 we have

$$\sum_{i=2}^{k+2} H(X_i) \geq 2k+1$$

and

$$\sum_{i=k+3}^{2k+3} H(X_i) \geq 2k+1.$$

Thus,

$$\sum_{i=1}^{2k+4} H(X_i) \geq 4k+4.$$

Hence,

$$\widetilde{\rho}^*(\mathcal{M}_k) \leq \frac{k+2}{2k+2}.$$

It is easy to see that the following complete multipartite covering Π of the graph $\mathcal{M}_k$ meets this bound.

$$\begin{aligned} \Pi = \Big\{ & \{X_1X_2, X_2X_3, \ldots, X_2X_{k+2}\}, \\ & \{X_3X_{k+3}, X_{k+3}X_{2k+3}\}, \\ & \vdots \\ & \{X_{k+2}X_{2k+2}, X_{2k+2}X_{2k+3}\}, \\ & \{X_{2k+3}X_{2k+4}\} \Big\}. \end{aligned}$$

□

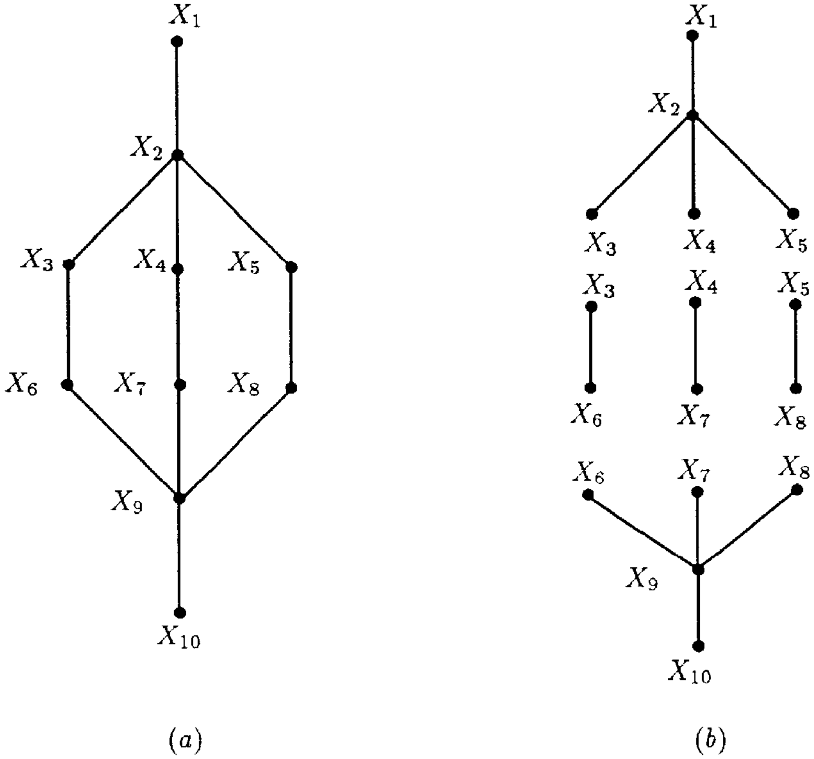


Figure 2

4.1 A $\mathcal{NP}$-completeness result

A close look to the proof of the upper bound in Theorem 8 shows that it can be applied also to any access structure $\mathcal{A}$ on $2k+2$ participants, $Y_0, X_0, X_1, \ldots, X_{2k}$, such that the set $\mathcal{A}$-*allowed* defined as

$$\mathcal{A}\text{-}allowed = \{Y_0X_0\}\bigcup\{X_0X_i, X_iX_{k+i} | 1 \leq i \leq k\}$$

is in the access structure, i.e. $\mathcal{A}$-*allowed* $\subseteq \mathcal{A}$, but the set $\mathcal{A}$-*forbidden* defined as

$$\mathcal{A}\text{-}forbidden = \{X_1X_2\ldots X_kY_0\}\bigcup\{Y_0X_{k+1}\}\bigcup\{X_1\ldots X_iY_0X_{k+i+1} | 1 \leq i \leq k-1\}$$

has no intersection with the access structure, i.e. $\mathcal{A}$-*forbidden* $\bigcap \mathcal{A} = \emptyset$. Let $\mathcal{B}_k$ be the set of all access structures which satisfy the above requirements. The sequence $(X_1, X_2, \ldots, X_k)$ is called the ***children list*** of access structure $\mathcal{A}$ (the name is inspired by the fact that the set $\mathcal{A}$-*allowed* has the form of a tree). To maintain simpler notation we denote a set $\{a_1, a_2, \ldots, a_n\}$ by the sequence $a_1a_2\ldots a_n$. In case the access structure is the closure of a graph, the set $\mathcal{A}$-*forbidden* can be written as

$$\mathcal{A}\text{-}forbidden\text{-}edges = \{Y_0X_i | 1 \leq i \leq 2k\}\bigcup\{X_iX_j | 1 \leq i < j \leq k\}$$

$$\bigcup\{X_iX_{k+j}|1 \leq i < j \leq k\}.$$

Let $\mathcal{A}$ be an access structure on a set P of participants. Given a subset of participants $P' \subseteq P$, we define the access structure *induced by* P' as the family of sets $\mathcal{A}[P'] = \{x \in \mathcal{A} | x \subseteq P'\}$. Extending Theorem 3.3 of [6] to general access structures and using Theorem 8 we can prove the following theorem.

Theorem 10. *Let $\mathcal{A}$ be an access structure on a set P of participants and $P' \subseteq P$. If $\mathcal{A}[P'] \in \mathcal{B}_k$, where $k \geq 1$, then the optimal information rates for $\mathcal{A}$ and $\mathcal{A}[P']$ satisfy*

$$\rho^*(\mathcal{A}) \leq \rho^*(\mathcal{A}[P']) \leq \frac{1}{2} + \frac{1}{4k+2},$$

and optimal average information rate for $\mathcal{A}[P']$ satisfies

$$\widetilde{\rho}^*(\mathcal{A}[P']) \leq \frac{2}{3} + \frac{2}{9k+6}.$$

Above theorem gives an upper bound on the information rate of access structures given that the access structure induced by a subset of participants is in $\mathcal{B}_k$. Unfortunately, testing for this property is an hard computational problem, as we show that this is NP-complete. Let $\mathcal{A}$ be an access structure, a set $C \in \mathcal{A}$ is a *minimal* set of $\mathcal{A}$ if it does not contain any set in $\mathcal{A} \setminus \{C\}$. Define the BOUNDED-INFORMATION-RATE problem as follows: Given a set of participants P and an access structure $\mathcal{A}$ defined by the family of minimal sets which can recover the secret and a positive integer k, determine if there is a subset $P' \subseteq P$ such that the induced access structure $\mathcal{A}[P']$ is in $\mathcal{B}_k$.

Theorem 11. *BOUNDED-INFORMATION-RATE is NP-complete.*

Proof. The proof will be given in the final version of the paper. □

4.2 Upper bounds for more general access structures

A general technique to upper bound the average information rate $\widetilde{\rho}^*(G)$, of graphs G who have one or more induced subgraphs of a given form is given below.

If G is a graph and $V_1 \subseteq V(G)$, then we define the *induced* graph $G[V_1]$ to have vertex set V_1 and edge set $\{XY \in E(G) : X, Y \in V_1\}$.

Let G be a graph. We define a subgraph F_G of G, that we will call the *foundation* of G, in the following manner. This is an extension of the notion of foundation presented in [4]. Let $X \in V(G)$. Let k be the maximum integer such that there is a set V' of $2k+1$ vertices $Y_0, X_1, \ldots, X_{2k} \in V(G)$ such that the induced subgraph $G[V' \cup \{X\}]$ is in $\mathcal{B}_k$; that is, $E(G[V' \cup \{X\}])$ contains the set $\mathcal{A}$*-allowed* but does not contain any edge in the set $\mathcal{A}$*-forbidden-edges.* Clearly $k < deg(X)$, where $deg(X)$ is the degree of vertex X. A set V' satisfying above properties is called a X*-set* of vertex X, with size k. Denote by $f_{X;X_1,\ldots,X_k}$ the set of edges XX_i, $i = 1, \ldots, k$. We call $f_{X;X_1,\ldots,X_k}$ the *local foundation* of vertex X and X-set V' and we call the vertices $X_1, \ldots, X_k$ *descendants* of X in $f_{X;X_1,\ldots X_k}$. Let $\{V_1, \ldots, V_{m_X}\}$ be the family of all X-sets of vertex $X \in V(G)$, and $\{f_X^1, \ldots, f_X^{m_X}\}$ be the family of the corresponding local foundations. Observe that this approach might not be feasible for large values of m, since m might be exponentially large in the worst case. Now we can define the *foundation* F_G of a graph G as follows

$$F_G = \{f_X^1, \ldots, f_X^{m_X} \,|X \in V(G)\}.$$

If $f_{X_0}^i$ is in F_G, the foundation of a graph G, and $X_1, \ldots, X_k$ are descendants of X_0 in $f_{X_0}^i$, then by Theorem 8, we have $\sum_{i=0}^{i=k} H(X_i) \geq (2k+1)H(S)$ for any secret sharing scheme with access structure $cl(E(G))$. Consider the following linear programming problem $\mathcal{A}(G)$:

Minimize $\qquad C = \sum_{X \in V(G)} a_X$

subject to:

$$a_X \geq 0, \quad X \in V(G)$$

$$a_{X_0} + \cdots + a_{X_k} \geq k\ , \quad X_0 \in V(G), \quad f_{X_0}^i \in F_G, \text{ and}$$

$$X_1, \ldots, X_k \text{ descendants of } X_0 \text{ in } f_{X_0}^i$$

The following upper bound on the average information rate holds.

Theorem 12. *Let G be a graph with foundation G_1. Let C^* be the optimal solution to the problem $\mathcal{A}(G)$. Then*

$$\tilde{\rho}^*(G) \leq \frac{|V(G)|}{C^* + |V(G)|}.$$

Proof. The proof will be given in the final version of the paper. □

5 Lower Bounds on Information Rate and Average Information Rate

In this section we will give several general lower bounds on the information rate and on the average information rate of access structures represented by graphs.

We first improve on the bound of Theorem 3 for graphs with n vertices and odd maximum degree d.

Lemma 13. *Let G be a graph of n vertices and maximum degree d, d odd. Then*

$$\rho^*(G) \geq \frac{1}{\lceil d/2 \rceil + 1 - \lceil d/2 \rceil / n}.$$

Proof. Let $Adj(X)$, $Inc(X)$, $degree\_one(X)$ be the following sets :

- $Adj(X) = \{Y : (X,Y) \in E\}$
- $Inc(X) = \{(X,Y) : (X,Y) \in E\}$
- $degree\_one(X) = \{Y \in Adj(X) : |Inc(Y)| = 1\}$

Let $X \in V(G)$ and G_x be a subgraph of G such that $V(G_x) = \{X\} \bigcup Adj(X)$ and $E(G_x) = Inc(X)$. It is well known a secret sharing scheme for G_x exists with information rate equal to 1 (G_x is a complete multipartite graph). Consider the graph G' where $V(G') = V(G) - \{X\} \bigcup degree\_one(X)$ and $E(G') = E(G) - Inc(X)$. We realize a secret sharing scheme for G', for a secret of one bit, using the technique showed in Theorem 3.8 of [BrSt]. Each vertex in $Adj(X) \bigcap V(G')$ gets at most $\lceil (d-1)/2 \rceil + 1$ bits while other vertices get at most $\lceil d/2 \rceil + 1$ bits. A secret sharing scheme for G can be realized joining the scheme for G_x and the scheme for G'. In this scheme the vertex X will receive one bit, the vertices in $Adj(X) \bigcap V(G')$ will receive at most $\lceil (d-1)/2 \rceil + 2$ bits, while other vertices will get at most $\lceil d/2 \rceil + 1$ bits. Since $\lceil (d-1)/2 \rceil + 2 = \lceil d/2 \rceil + 1$, if d is odd, there is a secret sharing schemes for G, for a secret consisting of a single bit, that gives to each vertex in G at most $\lceil d/2 \rceil + 1$ bits while a predeterminated vertex gets only one bit. If we consider n of these secret sharing schemes, one for each vertex in V, and then we compose them, we can realize a secret sharing scheme, for a secret of n bits, giving to each vertex at most $1 + (n-1)(\lceil d/2 \rceil + 1)$ bits, so we can realize a secret sharing scheme with an information rate equal to

$$\frac{1}{\lceil d/2 \rceil + 1 - \lceil d/2 \rceil / n},$$

and the lemma follows. □

For a graph G of maximum degree 3, the bound of [6] gives $\rho^*(G) \geq 1/3$ while the bound of lemma 13 gives $\rho^*(G) \geq 1/(3 - 2/n)$. The following lemma gives an improved bound.

Lemma 14. *Let G be a graph of maximum degree 3. Then, $\rho^*(G) \geq 2/5$.*

Proof. Consider a covering $\mathcal{C}$ of G consisting of maximal length paths $P_1, \ldots, P_m$. It is well know a secret sharing scheme for a path exists with an optimal information rate equal to 2/3 (see Theorem 5), this scheme, for a secret of two bits, gives two bits to terminal vertices in the path while other vertices gets three bits. We can realize a secret sharing scheme for G, for a secret of two bits, using secret sharing schemes, with optimal information rate, for the paths belonging to $\mathcal{C}$. A vertex of G of degree one can only be a terminal vertex of a path so it receive two bits. If a vertex has degree two then it belongs to only one path and it receives three bits, it cannot be a terminal vertex of two different paths since we consider a covering of maximal length paths. If a vertex has degree three then it can't belong to three different paths since we consider a covering of maximal length paths so it belongs to two paths, it is a terminal vertex of a path and it is a central vertex of another path and it gets totally five bits. Thus we can construct a secret sharing scheme for G, giving to each vertex at most five bits for a secret of two bits obtaining a secret sharing scheme with information rate equal to 2/5. □

If we know the number of vertices in the graph G then we can improve previous bound as stated by next lemma.

Lemma 15. *Let G a graph of maximum degree 3 with n vertices. Then,*

$$\rho^*(G) \geq \frac{2}{5 - 3/n}.$$

Proof. Let G_x, with $X \in V(G)$, be the graph defined in Lemma 13. Consider the graph G' where $V(G') = V(G) - \{X\} \bigcup degree\_one(X)$ and $E(G') = E(G) - Inc(X)$. We realize a secret sharing scheme for G', for a secret of two bit, using the technique showed in Lemma 14. Each vertex $Y \in Adj(X) \bigcap V(G')$ gets at most 3 bits, since $|Inc(Y)| \leq 2$, while the other vertices get at most 5 bits. A secret sharing scheme for G can be realized joining the scheme for G_x and the scheme for G'. Thus we can realize a secret sharing scheme for G, for a secret consisting of two bits, giving two bits to a predeterminated vertex while other vertices get at most five bits. If we consider n of these schemes, one for each vertex, and then we compose them we obtain a secret sharing scheme for a secret of $2n$ bits giving to each vertex at most $2 + 5(n-1) = 5n - 3$ bits so the information rate for this scheme is $2/(5 - 3/n)$. □

Applying the same reasoning of Lemma 14 to graphs of odd degree d leads to the bound $\rho^*(G) \geq 1/(\lfloor d/2 \rfloor 1.5 + 1)$ which is worse than previous bounds.

Regardless of the degree, it is possible to obtain better bounds for trees. We recall that an internal node is a vertex of degree greater than one.

Lemma 16. *Let G be a tree with n internal vertices. Then*

$$\rho^*(G) \geq \frac{n}{2n-1}.$$

Proof. In [4] was showed how to obtain a secret sharing scheme for any tree with information rate equal to 1/2. This scheme, for a secret consisting of a single bit, gives one bit to a predeterminated vertex $X \in V(G)$ and to all non-internal vertices, whereas each other vertex gets two bits. We will use this construction as basic construction. If we consider n of these schemes, one for each internal vertex, and we compose them then it is possible to realize a secret haring scheme for G, for a secret of n bits, giving to each vertex at most $2(n-1)+1 = 2n-1$ bits. Thus

$$\rho^*(G) \geq \frac{n}{2n-1}.$$

□

If only the number of vertices are known, what can we say on the information rate of a graph G? The maximum degree of G can be as bad as $n-1$. Thus, the bound of [6] gives $\rho^*(G) \geq 1/(\lceil (n-1)/2 \rceil + 1)$, while the bound of Lemma 13 gives $\rho^*(G) \geq 1/(\lceil (n-1)/2 \rceil + 1 - \lceil (n-1)/2 \rceil / n)$, if n is even.

In this last part of the paper we present general lower bounds on the information rate and average information rate for *any* graph G with n vertices. The lower bounds are obtained by using known results on the covering of the edges of a graphs by means of complete bipartite graphs. We first recall that Brickell and Davenport [5] proved that a graph G has information rate 1 if and only if G is complete multipartite graph.

Tuza [22] proved that the edge-set of an arbitrary graph G can be covered by complete bipartite subgraphs such that the sum of the number of the vertices of such subgraphs is less than $3n^2/2\log n + o(n^2/\log n)$. Using the above quoted result by Brickell and Davenport we get that the optimal average information rate for any graph G with n vertices is greater than n times the inverse of $3n^2/2\log n + f(n)$, where $|f(n)| < \epsilon n^2/\log n$, for all $\epsilon > 0$ and sufficiently large n. Therefore, the average information rate is greater than $2\log n/3n + g(n)$, where $|g(n)| \leq (2\epsilon/3(\epsilon + 3/2))\log n/n$, if $|f(n)| < \epsilon n^2/\log n$.

Feder and Motwani [10] proved that the problem of partitioning the edges of a graph G into complete bipartite graphs such that the sum of the cardinalities of their vertex sets is minimized is NP-complete. However, they proved that the edge set of a graph $G = (V, E)$, with $|V| = n$ and $|E| = m$ can be partitioned into complete bipartite graphs with sum of the cardinalities of their vertex sets $O(\frac{m \log \frac{n^2}{m}}{\log n})$, and presented an efficient algorithm to compute such a partition. Using their result, it follows that there is a secret sharing scheme with average information rate at least $\Omega(\frac{n \log n}{m \log \frac{n^2}{m}})$.

Finally, we recall a result of Erdös and Pyber [9] (see also [18]) which states that edges of a graph G with n vertices can be partitioned into complete bipartite graphs such that each vertex of G is contained by at most $O(n/\log n)$ complete bipartite graphs. This result directly implies that the optimal information rate of G is $\rho^*(G) = \Omega\left(\frac{\log n}{n}\right)$.

These results can be summarized in the following theorem.

Theorem 17. *Let G be a graph with n vertices and m edges. Then, the optimal average information rate for G satisfies*

$$\tilde{\rho}^*(G) > \frac{2\log n}{3n} + o\left(\frac{\log n}{n}\right),$$

and

$$\tilde{\rho}^*(G) = \Omega\left(\frac{n \log n}{m \log \frac{n^2}{m}}\right).$$

The optimal information rate for G satisfies

$$\rho^*(G) = \Omega\left(\frac{\log n}{n}\right).$$

It is worth pointing out that if G is a sparse graph, i.e., $m = \alpha n$, where α is a constant, then above theorem implies that $\tilde{\rho}^*(G)$ is limited from below by a constant. This result describes a wide class of graphs having average information rate that does not go to zero as the number of participants increases.

Acknowledgments

We are indebted with professor Capocelli for his constant encouragement and support. We would like to dedicate this paper to his memory as a sign of appreciation and love.

We would like thank L. Pyber for providing us reference [18] and A. Marchetti-Spaccamela and E. Feuerstein for bringing to our attention reference [10].

References

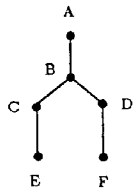
1. J. C. Benaloh and J. Leichter, *Generalized Secret Sharing and Monotone Functions*, in "Advances in Cryptology - CRYPTO 88", Ed. S. Goldwasser, vol. 403 of "Lecture Notes in Computer Science", Springer-Verlag, pp. 27–35.
2. G. R. Blakley, *Safeguarding Cryptographic Keys*, Proceedings AFIPS 1979 National Computer Conference, pp.313–317, June 1979.
3. C. Blundo *Secret Sharing Schemes for Access Structures based on Graphs*, Tesi di Laurea, University of Salerno, Italy, 1991, (in Italian).
4. C. Blundo, A. De Santis, D. R. Stinson, and U. Vaccaro, *Graph Decomposition and Secret Sharing Schemes*, Eurocrypt 1992, Hungary.
5. E. F. Brickell and D. M. Davenport, *On the classification of ideal secret sharing schemes*, J. Cryptology, 4:123–134, 1991.
6. E. F. Brickell and D. R. Stinson, *Some Improved Bounds on the Information Rate of Perfect Secret Sharing Schemes*, Lecture Notes in Computer Science, 537:242–252, 1991. To appear in J. Cryptology.
7. R. M. Capocelli, A. De Santis, L. Gargano, and U. Vaccaro, *On the Size of Shares for Secret Sharing Schemes*, in "Advances in Cryptology - CRYPTO 91", Ed. J. Feigenbaum, vol. 576 of "Lecture Notes in Computer Science", Springer-Verlag, pp. 101–113. To appear in J. Cryptology.
8. I. Csiszár and J. Körner, *Information Theory. Coding theorems for discrete memoryless systems*, Academic Press, 1981.
9. P. Erdős and L. Pyber, unpublished.
10. T. Feder and R. Motwani, *Clique Partition, Graph Compression and Speeding-up Algorithms*, Proceedings of the 23*rd* Annual ACM Symposium on Theory of Computing, New Orleans, 1991, pp. 123–133.
11. R. G. Gallager, *Information Theory and Reliable Communications*, John Wiley & Sons, New York, NY, 1968.
12. M. Garey and D. Johnson, *Computers and Intractability: a Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., New York, 1979.
13. E. D. Karnin, J. W. Greene, and M. E. Hellman, *On Secret Sharing Systems*, IEEE Trans. on Inform. Theory, vol. IT-29, no. 1, Jan. 1983, pp. 35–41.
14. S. C. Kothari, *Generalized Linear Threshold Schemes*, in "Advances in Cryptology - CRYPTO 84", G. R. Blakley and D. Chaum Eds., vol 196 of "Lecture Notes in Computer Science", Springer-Verlag, pp. 231–241.

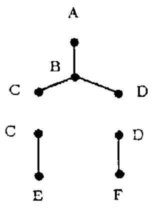
15. M. Ito, A. Saito, and T. Nishizeki, *Secret Sharing Scheme Realizing General Access Structure,* Proc. IEEE Global Telecommunications Conf., Globecom 87, Tokyo, Japan, 1987.
16. K. M. Martin, *Discrete Structures in the Theory of Secret Sharing,* PhD Thesis, University of London, 1991.
17. K. M. Martin, *New secret sharing schemes from old,* submitted to Journal of Combin. Math. and Combin. Comput..
18. L. Pyber, *Covering the Edges of a Graph by ...,* in Sets, Graphs and Numbers, Colloquia Mathematica Soc. János Bolyai, L. Lovász, D. Miklós, T. Szönyi, Eds., (to appear).
19. A. Shamir, *How to Share a Secret,* Communications of the ACM, vol. 22, n. 11, pp. 612–613, Nov. 1979.
20. G. J. Simmons, *An Introduction to Shared Secret and/or Shared Control Schemes and Their Application,* Contemporary Cryptology, IEEE Press, pp. 441–497, 1991.
21. D. R. Stinson, *An Explication of Secret Sharing Schemes,* Technical Report UNL-CSE-92-004, Department of Computer Science and Engineering, University of Nebraska, February 1992.
22. Z. Tuza, *Covering of Graphs by Complete Bipartite Subgraphs; Complexity of 0-1 matrices,* Combinatorica, vol. 4, n. 1, pp. 111–116, 1984.

Appendix A

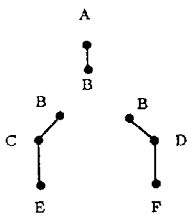
In this appendix we analyze all graphs who have optimal information rate less than 2/3 accordingly to Theorem 10. The schemes for these graphs are obtained by using the Multiple Construction Technique [4] based on complete multipartite coverings of the graph. The optimal information rate is not greater than 3/5 and the optimal average information rate is less than or equal to 3/4 for all graphs from Theorem 10. All these results are summarized in Table 1, and the first CMC of each graph gives the scheme with average information rate showed in Table 1. Below are depicted some of the minimal CMCs for 5 graphs on 6 vertices.

G_1



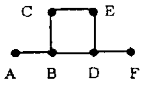


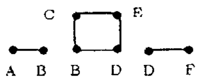
Π_1



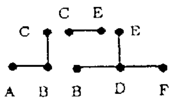
Π_2

G_2

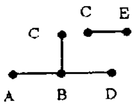




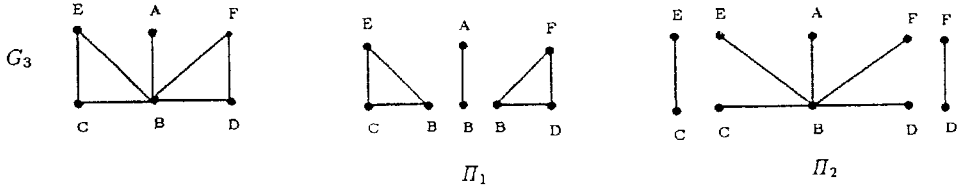
Π_1

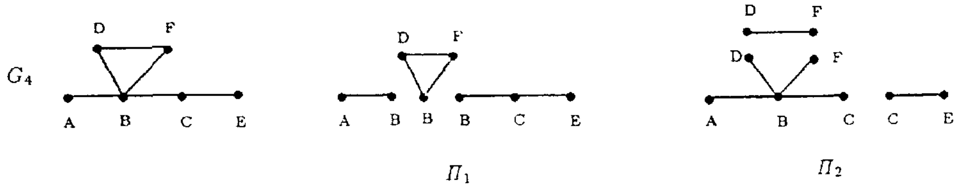


Π_2



Π_3





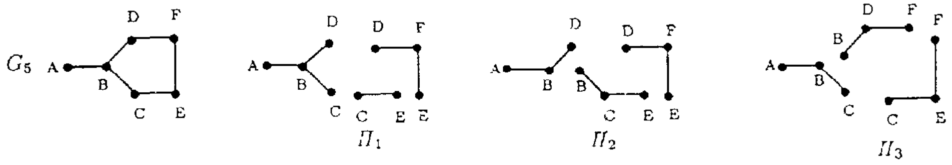


Table 1. Information Rate and Average Information Rate

Graph	Information Rate	Average information Rate
G_1, G_2, G_3, G_4	$\rho^* = 3/5$	$\widetilde{\rho}^* = 3/4$
G_5	$\rho^* = 3/5$	$2/3 \leq \widetilde{\rho}^* \leq 3/4$