Computer Science Journal of Moldova, vol.**2**, no.1(4), 1994

A new approach to inforamtion coding and protection based on the theory of matroids

V.Borshevich, W.Oleinik

Abstract

A new aproach to coding and protection of information in the computer and telecommunication systems is proposed and discussed. It is based on the mathematical apparatus of the theory of matroids, which in combination with the randomization method gives the possibility to protect large volumes of information against spoils, having considerably high speed of coding/decoding of information even on the PC platform. The proposed approach open the way for building of a new class of codes usable for recovering of large amount of information with high degree of accuracy.

The proposed approach to information coding and protection consists in using of a number of remarkable properties of matroids [1] as abstract structures of independency. As elements of matriod can be used vectors, graphs, geometric objects etc. The interpretation of the arrays of data in the storage and communication systems as such kind of objects opens the way to new methods of data coding and protection.

Let to discuss next basic mathematical model used in our approach. We define some matroid as a set of S elements and the closure statement $^{-}$ ([1]):

$$\mathcal{M}(S) = \langle S, ^{-} \rangle, \tag{1}$$

where

$$S = B_o \cup C, \tag{2}$$

B_o — is the distinct basis of matroid,

$$C: \; C \subset \overline{B}_o, \qquad C \cap B_o = \emptyset \tag{3}$$

The set C is named redundant against to basis B_o, which will corresponds to information under protection of volume $|B_o| = n$. If we will take into account, that one element of matroid is interpreted as a bulk of information of k bytes (for example, as one sector of $k = 512$ bytes), then the initial information will have the volume of $n \cdot k$ bytes and the redundancy of $r \cdot k$ bytes, where $r = |C|$.

If the information about some subset $\varepsilon_s : \ \varepsilon_s \subseteq S$, of elements of matroid was lost or spoiled during transmitting or storaging of data, then in the possesion of user remains the submatroid

$$\mathcal{M}(S_e) = \langle S_e, \overline{\ } \rangle, \tag{4}$$

where

$$S_e = S \setminus \varepsilon_s$$

is the set of remainder elements from initial set S. It is easy to note, that if the set S_e contains even one basis, then the initial matroid $\langle S, \overline{\ } \rangle$ and its (which is main) the distinct basis B_o can be recovered.

The most important from this point of view are uniform matroids, which are characterized by the next property: every subset of cardinal number equal to the cardinal number of basis is basis. This mean that having losses ε_s of cardinal number

$$|\varepsilon_s| \leq |S| - n \tag{5}$$

we can recover all initial information. The condition (5) can be presented in another way:

$$|\varepsilon_e| \leq r, \tag{6}$$

from where folows, that it is permissible to lose up to $r \cdot k$ bytes and to recover all initial information of $n \cdot k$ bytes.

However the uniformity can not be achieved for any class of matroids. For example it can be shown that in the class of binary matroids, which presents especialy high practical interest, a free matroid can be obtained only for the case of $r = 1$. Therefore, it can be found such situations in which the conditions (5)–(6) will not guarantee the full recovering of initial information.

From the point of view of the games theory [2] such kind of situation can be considered as a game with the nature having a set of the strategies $Y = \{\varepsilon_s : |\varepsilon_s| \leq r\}$ with the gain function

$$\mathcal{H}(\varepsilon_s, S) = \begin{cases} 1, & \text{is} \quad S \setminus \varepsilon_s \quad \text{contines at least one basis} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

If is clear that a minimax solution of the clear strategies exists only for the subset of the nature strategies of the next form $Y_l = \{\varepsilon_s : |\varepsilon_s| \leq l < r\}$, where l can be unacceptable lower then r, and this mean, that the solution must be searched in the domain of mixed strategies, in other words in the probabilistic strategies area.

From the practical point of view this suppose the introduction of the randomization, or the random reflections like next:

$$\langle \varphi, p(\varphi) \rangle, \quad \varphi \in \Phi \tag{8}$$

where

$$\varphi : \quad V \longrightarrow S, \tag{9}$$

V is a set of memory blocks (sectors, clusters etc.) containing the information about the elements of matroid $\mathcal{M}(S)$; φ — is the reflection from V into S, which establish the distribution of the information about elements betweem the memory blocks; Φ — is the permisible set of such reflections, $p(\varphi)$ — is the probability of selection of φ from Φ.

In this case if the bulk of bassed blocks of memory ε_v, $\varepsilon_v \subseteq V$ is fixed the probability of losses ε_s from S is defined by the next value:

$$P\{\varepsilon_s = \varphi(\varepsilon_v)\} = \sum_{\varphi : \varphi(\varepsilon_v) = \varepsilon_s} p(\varphi) \tag{10}$$

and the probability of full recovery of information — by the next value:

$$Q = \sum_{\varepsilon_s : \mathcal{H}(\varepsilon_s, S) = 1} p(\varepsilon_s). \tag{11}$$

The dependency $Q = Q(S, |\varepsilon_v|)$ is the main relation, which characterize the quality of data protection.

An experimental data protection system (data archiver) was elaborated by authors of actual paper based on the methods of the matroid's theory and the randomization. The system show high efficiency of proposed approach. For example if the data has $n \cdot k$ equal to about 1 Mbyte and the redudancy $r/n = 0.2$ it is guaranted full recovery of the information if lossins not exceeds 0.2 Mbyte with the probability $Q > 0.9995$. Using precompression of data do not produces enlargements of the initial data volume.

The significant particularity of the proposed approach consist in the exclusive high speed of coding/decoding processes for the large amount of information, especialy when binary matroids are used.

References

[1] M.Aigner. Cobinatorial Theory, Springer–Verlag, Berlin, Heidelberg, New York, 1979.

[2] H.Kuhn. Extensive Games and the Problem of Information, Contributions to the Theory of Games, vol II, Princeton, 1953, pp.193–216.

V.Borshevich, W.Oleinik
Technical University of Moldova,
Str. Stefan cel Mare, 168
Kishinev, 277012, Moldova
phone: (373–2) 23–76–13

Received 27 January, 1994