

# **INTERVAL MODEL OF MULTI-CRITERION TASK OF REENGINEERING PHYSICAL STRUCTURES OF DISTRIBUTED DATABASES**

Beskorovainyi V., Kolesnyk L.

*An interval mathematical model of a multicriteria task of reengineering physical structures of distributed databases is proposed. The model is built on the basis of the Kolmogorov-Gabor polynomial and the universal utility function of partial criteria. It provides an assessment of options for interval values of cost indicators, access time and network traffic. The use of the universal function of general utility makes it possible to more accurately take into account the preferences of the decision-maker. The use of a rational S-like utility function of partial criteria allows to reduce the time complexity of the procedures for calculating estimates. Practical application of the proposed model will improve the efficiency of automated design technologies for distributed databases.*

## **INTRODUCTION**

Distributed databases (DDBs) are becoming the most important elements of modern control systems, monitoring systems, CADs systems, and other objects from various spheres of human activity. In modern companies, geographically distributed data processing centers are created, consisting of a large number of computers, united in high-speed networks containing both distributed and parallel systems [1]. The volumes of data stored in distributed environments tend to grow rapidly, which, regardless of the principle of building control systems (fragmentation or replication), leads to the problem of managing them. At the same time, the costs of creating and the efficiency of DDB functioning largely depend on their physical structures, which determine the physical implementation of the information structure of databases on computer networks [2].

Changes in technologies, element base, operating conditions or functional requirements for DDB at a certain stage lead to the need to reengineer their

infrastructure, which is based on information and computer networks. At the same time, DDB (as geographically distributed objects) is characterized by the dependence of their functional and cost characteristics on the topology of local bases (LB) [3]. This requires solving a complex of combinatorial optimization problems taking into account a variety of functional and cost indicators.

Modern technologies for the design of physical structures DDB are based on the theory of aggregate-decomposition and block-hierarchical approaches. Their use assumes the division of DDB descriptions as a design object according to the degree of detail into hierarchical levels and aspects, and the design process into groups of procedures associated with obtaining and transforming descriptions with respect to the selected levels and aspects. In the future, the obtained descriptions in the form of mathematical models are aggregated to obtain solutions for the project as a whole.

The main theoretical results were obtained for solving partial one-criterion problems of the synthesis of physical structures of DDB, taking into account various types and architectures of databases and computer networks, partitioning strategies and data access protocols under conditions of definite goals and data [2]. Taking this into account, the scientific and applied problem of increasing the efficiency of DDB computer-aided design technologies by developing multi-criteria mathematical models for the synthesis of their physical structures under conditions of incomplete definition of goals and data is urgent.

## **1. BASIC MODEL OF MULTI-CRITERION TASK OF REENGINEERING PHYSICAL STRUCTURES OF DISTRIBUTED DATABASES**

The task of reengineering the physical structures of DDB is considered in the following setting. Given: a set of users of a distributed base  $I = \{i\}$ ,  $i = \overline{1, n}$ , connected by a computer network  $G = (I, R)$  (where  $R = [r_{ij}]$ ,  $i, j = \overline{1, n}$  is an adjacency matrix that defines a set of direct connections between elements and nodes of a computer network); many information resources (IR) in the form of files, funds

and  $J = \{ j \}$ ,  $j = \overline{1, m}$ ; IR volumes are  $l = [ l_j ]$ ,  $j = \overline{1, m}$ ; the intensity of incoming requests from each node of the computer network to each of the IR are  $\lambda = [ \lambda_{ij} ]$ ,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ ; the volumes of requests to the IR from the nodes of the computer network  $a = [ a_{ij} ]$ ,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ ; costs of storing IR in the nodes of a computer network  $c(x) = [ c_{ij}(x) ]$   $i = \overline{1, n}$ ,  $j = \overline{1, m}$ ; ( $x = [ x_{ij} ]$  matrix displaying the placement of IRs on the nodes of a computer network;  $x_{ij} = 1$  if the  $j$ -th IR is stored in the  $i$ -th network node;  $x_{ij} = 0$  - otherwise);  $c_t$  - the cost of transferring a unit of information;  $d = [ d_{ij} ]$  - the amount of information transmitted when updating the IR from network nodes;  $z = [ z_{ij} ]$  - matrix of IR updates ( $z_{ij} = 1$  if the  $j$ -th IR is updated from the  $i$ -th node of the computer network;  $z_{ij} = 0$  - otherwise);  $X = \{ x \}$  - the set of valid implementations of the physical structure of the DDB.

It is necessary to determine the best option of the physical structure of the DDB from the many permissible  $x^o \in X$  (the number of local bases (LB), the distribution of IR over the LB  $x = [ x_{ij} ]$ , the placement of LB over the network nodes, the amount of storage devices for storing LB, the throughput of channels between elements and nodes of the computer network.

Optimization criteria:

– the cost of  $x$  - implementation of the physical structure of DDB:

$$c(x) = \sum_{i=1}^n \sum_{j=1}^m c_{ij}(x) x_{ij} + c_t \sum_{i=1}^n \sum_{j=1}^m (\alpha_{ij} + \beta_{ij}) x_{ij} + c_t \sum_{i=1}^n \sum_{j=1}^m d_{ij} z_{ij} \rightarrow \min_{x \in X}, \quad (1)$$

where  $\alpha_{ij} = \lambda_{ij} a_{ij} T$  – the total volume of requests to the  $j$ -th IR from the  $i$ -th node of the computer network ( $\lambda_{ij}$  and  $a_{ij}$  – the intensity of requests and the volume of a request to the  $j$ -th IR from the  $i$ -th node of the network;  $T$  – time interval over which cost estimates are made),  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ ;  $\beta_{ij}$  – the total volume of responses to requests to the  $j$ -th IR from the  $i$ -th node of the network,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ ;

– access time to information resources in the  $x$ -implementation of the physical structure of DDB:

$$t(x) = \frac{1}{n} \sum_{i=1}^n [t_{ij}^{tr}(x) + t_{ij}^{pr}(x) + t_{ij}^{qp}(x) + t_{ij}^{rp}(x)] x_{ij} \rightarrow \min_{x \in X}, \quad (2)$$

where  $t_{ij}^{tr}(x)$  – time of request transmission from the  $i$ -th node to the  $j$ -th IR,  $i = \overline{1, n}$ ,  $j = \overline{1, m}$ ;  $t_{ij}^{pr}(x)$  – waiting time in the queue for a request from the  $i$ -th node by the  $j$ -th IR;  $t_{ij}^{qp}(x)$  – processing time of a request from the  $i$ -th node by the  $j$ -th IR;  $t_{ij}^{rp}(x)$  – time of transmission of a response to a request from the  $i$ -th node via the  $j$ -th IR;

– volume of transmitted information at  $x$ -implementation of the physical structure:

$$v(x) = \sum_{i=1}^n \sum_{j=1}^m (\alpha_{ij} + \beta_{ij}) x_{ij} + \sum_{i=1}^n \sum_{j=1}^m l_{ij} z_{ij} \rightarrow \min_{x \in X}. \quad (3)$$

At the same time, it is required to ensure the completeness of the database by distributing all IRs (with possible duplication) over local databases  $\sum_{i=1}^n x_{ij} \geq 1$   $\forall j = \overline{1, n}$  and fulfilling the restriction on the time of access to the information resources of the database  $t(x) \leq t^*$  (where  $t^*$  is the maximum admissible value of the access time to the IR).

In the process of design or reengineering, the costs of implementing the physical structure (1), the time of access to information resources (2) and the amount of transmitted information (3) for each of the implementations of the physical structure of DDB  $x \in X$  are determined with some error, which must be taken into account when choosing the final solution. The use of the apparatus of fuzzy mathematics [4] to take into account the errors of the calculated estimates  $c(x)$ ,  $t(x)$  and  $v(x)$  in the procedures of automatic generation and estimation, due to the high time complexity, is ineffective. For this purpose, we use the apparatus of interval

analysis [5-8], which allows obtaining solutions of sufficient accuracy with much lower computational costs.

## 2. INTERVAL ESTIMATION OF CHARACTERISTICS OF PHYSICAL STRUCTURES DDB

In the proposed basic model of the multi-criteria problem of reengineering of physical structures DDB, we use the interval representation of characteristics  $c(x) = [c^-(x); c^+(x)]$ ,  $t(x) = [t^-(x); t^+(x)]$  and  $v(x) = [v^-(x); v^+(x)]$  of the options  $x \in X$ . In this case, each of the characteristics will be represented not by one machine number, but by two that define its boundaries. For some values of characteristics  $a$ ,  $b$  and  $c$  the rules for performing operations of classical interval arithmetic [5]: "If  $[c^-; c^+] = [a^-; a^+] \circ [b^-; b^+]$ ,  $a \in [a^-; a^+]$ ,  $b \in [b^-; b^+]$  then  $a \circ b = [c^-; c^+]$ , where  $\circ$  is the symbol of the arithmetic operation from the set  $\{+, -, *, /\}$ ". In this case, the results of operations of addition, subtraction, division and multiplication over intervals are given by the relations:

$$\begin{aligned}
 [a] + [b] &= [a^- + b^-; a^+ + b^+], \\
 [a] - [b] &= [a^- - b^+; a^+ - b^-], \\
 [a] \cdot [b] &= [\min\{a^- \cdot b^-, a^- \cdot b^+, a^+ \cdot b^-, a^+ \cdot b^+\}; \max\{a^- \cdot b^-, a^- \cdot b^+, a^+ \cdot b^-, a^+ \cdot b^+\}], \\
 [a] / [b] &= [a] \cdot [1/b^+; 1/b^-].
 \end{aligned} \tag{4}$$

The main problems of classical interval arithmetic are solved in extended interval arithmetic [7, 8]. It is known that the width of the resulting interval in interval calculations depends on the order of operations.

The basis of numerical methods implemented in interval arithmetic is a property called "monotonicity of inclusion" [5].

Let us have intervals  $[a^-; a^+]$ ,  $[b^-; b^+]$ ,  $[c^-; c^+]$ ,  $[d^-; d^+]$ . Then the

following feature of monotonicity of the inclusion is confirmed:

$$\begin{aligned} & [a^-; a^+] \subset [c^-; c^+], [b^-; b^+] \subseteq [d^-; d^+] \Rightarrow \\ & \Rightarrow [a^-; a^+] \cap [b^-; b^+] \subseteq [c^-; c^+] \cap [d^-; d^+]. \end{aligned} \quad (5)$$

This feature allows to carry out iterative procedures for approximating the sets of solutions for tasks with interval (imprecise) data. As a result, the cardinality of the sets is determined by the width of the input data intervals and the capabilities of the iterative procedures [6].

### 3. FORMATION OF A GENERALIZED CRITERION FOR EVALUATING THE EFFECTIVENESS OF DECISIONS

The combinatorial nature of the tasks of reengineering the physical structures of DDB involves the use of automatic quantitative evaluation of options from many acceptable  $x \in X$ . A utility theory is used as a methodological basis for constructing estimates, according to which the value of its usefulness  $P(x)$  can be determined for each of the alternatives  $x \in X$  [9, 10]. At the same time for all  $x, y \in X$ :

$$x \sim y \leftrightarrow P(x) = P(y); \quad x \succ y \leftrightarrow P(x) > P(y); \quad x \succeq y \leftrightarrow P(x) \geq P(y). \quad (6)$$

Formally, the task of choosing the best option for a DDB physical structure  $x^o \in X$  can be reduced to the task of optimizing of the type:

$$x^o = \arg \max_{x \in X} P(x). \quad (7)$$

Let us introduce the notation for the selected optimization criteria:

$$k_1(x) = c(x) \rightarrow \min_{x \in X}, \quad k_2(x) = t(x) \rightarrow \min_{x \in X}, \quad k_3(x) = v(x) \rightarrow \min_{x \in X}. \quad (8)$$

For scalar estimation of options, it is proposed to use a universal utility function built on the basis of the Kolmogorov-Gabor polynomial [11], which, taking into account the selected set of partial criteria, will have the form:

$$P(x) = \sum_{i=1}^3 \lambda_i \xi_i(x) + \sum_{i=1}^3 \sum_{j=i}^3 \lambda_{ij} \xi_i(x) \xi_j(x) + \sum_{i=1}^3 \sum_{j=i}^3 \sum_{l=j}^3 \lambda_{ijl} \xi_i(x) \xi_j(x) \xi_l(x), \quad (9)$$

where  $\lambda_i, \lambda_{ij}, \lambda_{ijl}$  – weight coefficients of partial criteria  $k_i(x)$ ,  $i = \overline{1,3}$  and their

products  $\lambda_i \geq 0, \lambda_{ij} \geq 0, \lambda_{ijl} \geq 0$ ;  $\xi_i(x)$  – utility function of the partial criterion  $k_i(x)$ ,  $i = \overline{1,3}$ .

Quantitative evaluation of weight coefficients of partial criteria can be carried out by classical expert methods [8] or using the technology of comparator identification [10, 12]. To narrow the set of search solutions to the task, it is advisable to pre-isolate or parallel with generation the formation of a subset of effective (Pareto-optimal) variants.

To assess the utility of the values of partial criteria, we use the glue function, which has advantages in terms of accuracy and the number of computer operations to calculate it by comparison with the Gauss, Harrington and logistic functions [11]:

$$\xi_i(x) = \begin{cases} \overline{a}_i \cdot (b_{i1} + 1) \cdot \left\{ 1 - \left[ b_{i1} / \left( b_{i1} + \frac{\overline{k}_i(x)}{\overline{k}_{ia}} \right) \right] \right\}, \\ 0 \leq \overline{k}_i(x) \leq \overline{k}_{ia}; \\ \overline{a}_i + (1 - \overline{a}_i) \cdot (b_{i2} + 1) \cdot \left\{ 1 - \left[ b_{i2} / \left( b_{i2} + \frac{\overline{k}_i(x) - \overline{k}_{ia}}{1 - \overline{k}_{ia}} \right) \right] \right\}, \\ \overline{k}_{ia} < \overline{k}_i(x) \leq 1, \end{cases} \quad (10)$$

where  $\overline{k}_{ia}, \overline{a}_i$  – normalized values of function component gluing point coordinates,  $0 \leq \overline{k}_{ia} \leq 1, 0 \leq \overline{a}_i \leq 1$ ;  $b_{i1}, b_{i2}$  – parameters, nonlinearities of the initial and final components of the function.

## CONCLUSIONS

An interval mathematical model of the task of reengineering physical structures of DDB based on universal functions of general utility and utility of partial criteria is proposed. It allows to make a generalized assessment of options taking into account the indicators of cost, speed and network traffic in the presence of interval uncertainty of functional and cost characteristics. The use of universal utility functions allows more accurately taking into account the preferences of the decision

maker and reducing the time complexity of evaluation calculation procedures. The direction of further research may be the development of effective methods for optimizing the physical structures of DDB in conditions of incomplete certainty of goals and data.

## REFERENCES

1. Özsu, M.T., Valduriez, P. (2020), *Principles of Distributed Database Systems Fourth Edition*, New York: Springer, 674 p. DOI: 10.1007/978-3-030-26253-2
2. Beskorovainyi, V.V., Ulianova, O.S. (2010), "Mathematical models of multicriteria synthesis of physical structures of distributed databases", *Eastern-European Journal of Enterprise Technologies*, Vol. 4/4(46), P. 44–48 (in Russian).
3. Beskorovainyi, V.V., Podoliaka, K.E. (2015), "Development of systemological model of the problem of structural and topological reengineering of large-scale monitoring system", *Eastern-European Journal of Enterprise Technologies*, Vol. 3(75), P. 37–42 (in Russian).
4. Raskin L.G., Seraja O.V. (2008), *Fuzzy Mathematics. Fundamentals of the theory*. Kharkiv: Parus, 352 P.
5. Kalmykov, S.A., Shokin, Y.I., Iuldashev, Z.K. (1986), *Interval analysis methods*, Novosibirsk: Nauka, 222 p. (in Russian).
6. Dyvak, M.P. (2011), *Problems of mathematical modeling of static systems with interval data*, Ternopil: Vidavnytstvo TNEU "Ekonomichna dumka", 216 p. (in Ukraine)
7. Yakovlev, A.G. (1986), "Machine arithmetic of multi-intervals", *Problems of Cybernetics*, Vol. 125, p. 66–81. (in Russian)
8. Dimitrova, N.S., Markov, S.M., Popova, E.D., Atanassova, L., Herzberger J. eds. (1992), "Extended interval arithmetics: new results and applications", *Computer Arithmetic and Enclosure Methods*, Amsterdam: North-Holland, P. 225–232.
9. Kruchkovskiy, V.V., Petrov, E.G., Sokolova, N.A., Hodakov, V.E. (2013), "Introduction to the normative theory of decision-making", Herson: Grin D.S., 284 p. (in Russian).
10. Petrov, E, Brynza, N, Kolesnyk, L, Pisklakova, O. (2014), "Methods and models of decision-making under conditions of multi-criteria and uncertainty", Herson: Grin D.S., 192 p.
11. Beskorovainyi, V., Berezovskiy, H. (2017), "Identification of preferences in decision support systems", *ECONTECHMOD*, Vol. 06, No. 4, P. 15–20.
12. Ovezgeldiev, A.O., Petrov, E.G., Petrov, K.E. (2002), "Synthesis and identification of multifactorial models for estimation and optimization", Kiev: Naukova dumka, 164 p. (in Russian).