Modelling and optimization of information systems with distributed databases

I Ya Lvovich¹, Ya E Lvovich², A P Preobrazhenskiy¹, Yu P Preobrazhenskiy¹ and O N Choporov²

¹Information Systems and Technologies Department, Voronezh institute of high technologies, 73a, Lenina st, Voronezh, 394043, Russia
²Information Security Department, Voronezh state technical university, 14, Moscow dist., Voronezh, 394026, Russia

E-mail: komkovvivt@yandex.ru

Abstract. The paper deals with the problems of modern corporate information systems. It is necessary to carry out the structural synthesis of their components. It is important that there is a choice of the required number of parallel processing in the databases, as well as the assignment of priorities for the execution of the corresponding processing conditions. On the basis of enumeration schemes, alternative variables will be introduced. When calculations are carried out using the developed models, it is possible to calculate the values of predictive estimates for each of the search steps. That is, many promising options for information and telecommunication systems are selected, with the additional inclusion of procedures for taking into account restrictions and multi-criteria.

1. Introduction

When designing wide classes of corporate information systems (CIS), it is necessary to correctly select the components that make up their composition. One of the central components is the database. Recently, they have been designed in a distributed manner. It is required to solve problems associated with the structural synthesis of such components. Why is this necessary? When implementing the synthesis procedures, it is necessary to carry out work on the construction of a multi-alternative optimization model. On its basis, it is possible to build interesting predictive assessments from the point of view of practice. There are some peculiarities during their construction. They are connected with the fact that when the objective function is formed in the optimization problem, then its values that were determined for each of the search steps in the calculated predictive estimates determine the need to implement a simulation experiment using transactional calculations. This makes it possible to refine the characteristics of the analyzed information systems [1].

These models have specific features. This leads to the fact that modified algorithms are used for selecting a set of promising options for information and telecommunication systems and additional inclusion of procedures for taking into account restrictions and multicriteria. There may be several options. If a multicriteria choice is carried out according to a rational option, then the process of combining assessment according to fuzzy requirements and a randomized search for a compromise will
take place [2].

2. Characteristics of multi-alternative optimization based on a simulation model of a complex information and telecommunications system

When we carry out optimization using procedures for structural synthesis of such a component of the CIS as a distributed database, then it is important to solve the following problem. On different workstations, we can see some parts of the distributed database. It is part of the larger CIS. In a large corporate information system, there are sections of the local computer network. It can be divided into many computers $M = \{ M_i \}, i = 1, I$, which are considered as having similar characteristics [3]. Many files $F = \{ f_j \}, j = 1, J$ are contained within a distributed database. Interaction within the analyzed information and telecommunication systems occurs using network equipment [4].

It is important to distribute the database files among the components of information and telecommunication systems. This should be done in such a way as to minimize the processing time of requests. How can you ensure this minimization? An optimization model is being built. To create it, we introduce alternative variables:

$$
\begin{align*}
    x_{ji} &= \begin{cases} 
    1, & \text{if } j \text{-th file is on } i \text{-th computer}, \\
    0, & \text{otherwise,}
    \end{cases} \quad j = 1, J, \quad i = 1, I. 
\end{align*}
$$

Complex processes are taking place in the information and telecommunication system. There is a dependence of the processing time of requests on the corresponding values of the variables $x_{ji}, j = 1, J, i = 1, I$. To build it, we need to use simulation modeling. The model will be like this $T = \Psi(\{x_{ji}\})$. The objective function should be minimized. Appropriate restrictions are created for it:

$$
\begin{align*}
    x_{ji} &= \begin{cases} 
    1, & j = 1, J, \ i = 1, I. 
    \end{cases}
\end{align*}
$$

1) For any file $f_1$ with data, only one of the analyzed computers can be associated at the moment

$$i = 1, \sum_{j=1}^{I} x_{ji} = 1, \quad j = 1, J;$$

2) There are limitations for a computer in terms of the total amount of information that will be for it ... It is bounded from above by the value $V_i$

$$\sum_{j=1}^{J} v_j x_{ji} \leq V_i, \quad i = 1, I$$

here $V_i$ - it will be the amount of information inside the $j$-th file.

After taking into account these constraints, we can create a multi-alternative optimization model. On its basis, a structural synthesis of databases takes place [5, 6], which will be used inside information and
telecommunication systems

\[ T = \Psi(x_{ji}) \rightarrow \min, \quad \sum_{i=1}^{J} x_{ji} = 1, \quad j = 1, J, \]

\[ \sum_{j=1}^{J} v_{j} x_{ji} \leq V_{i}, \quad i = 1, I, \quad x_{ji} = \begin{cases} 1, & j = 1, J \\ 0, & j = 1, J \end{cases}, \quad i = 1, I. \quad (4) \]

It is required for practical applications to determine the dependency \( T = \Psi(x_{ji}) \). This can be done, for example, based on a transaction hierarchy approach. In this case, a simulation model is applied.

Interconnected structures should be formed in the information and telecommunication system. For this purpose, modeling of the CIS architecture is applied. Modeling transactions by architectural features must be carried out in the correct way. Therefore, it is necessary to structure structures at the architectural level when designing, it is required to correlate with various types of links.

When transactional computations are organized [7], then objects will be considered in the form of resources, correlated with the transactions being carried out. Then we can identify key links between transport and logical structures (figure 1):

- resources can be shared using transactions [8].
- object methods will influence how transaction bodies will be formed.
- if objects interact, then integrity constraints will be allocated for them using transactions.

![Figure 1. Illustration of the relationship between logical and transactional structures.](image)

It is important to note that the flow of transactions in relation to objects acts like a work flow. What are the features of this? There are still additional constraints on the integrity of the computations performed [9].

Components need to be labeled appropriately. We consider control flows as components in the process structure. In this regard, transactions show how the workflows will occur. At the same time, we support the condition that integrity will be ensured in the calculations that are performed.

Modular and transactional structures will be linked by means of objects. In this case, the transition from one holistic state to another will be within the framework of the transaction.

The required modular structure depends on what work should be done in order for the transaction to be tested appropriately [10].

Due to the physical structure of the information component, telecommunication systems will be
mapped accordingly to the hardware components.

What needs to be done in order to correctly describe the components of transactional structures?

It is required that links with existing workflows, particularities of hierarchies in transactions, characteristics of related objects, how the objects will be associated, in which way will be to a consistent state using the appropriate sequence of transactions. To use the hierarchy of transactions that is required to support organizational and structural relationships among transactions, it is necessary to use class diagrams. As a result, based on the above components, there are opportunities to build a simulation model. On its basis, key parameters in the designed information and telecommunication systems can be shown. Modeling of a set of options for information and telecommunication systems is carried out within the framework of simulation. The results obtained can be used in order to obtain a multi-criteria assessment on them. The next step is how decisions will be made [11].

3. Carrying out a multi-criteria choice according to options subject to unclear requirements

We have several options for information and telecommunication systems. To make a choice on them, you can use a method in which there is a multi-criteria choice for alternatives of options. In this case, a compositional rule is used, which allows aggregation by descriptions of alternatives. We proceed from the assumption that information about the preferences of persons who will make decisions is represented as fuzzy judgments. The main problem that will be in multi-criteria choice when fuzzy models are applied is how the assessments and related criteria. Why does this problem arise? This is due to the fact that different results will be obtained for the same problems, but using different methods. To give a solution to such problems, you can use a heuristic approach. We believe that many solutions [12] can be described using a set of criteria \( x_1, \ldots, x_p \). Then we say that linguistic variables are applied. They will be set on the basis of base sets \( U_t, \ldots, U_p \). Let’s give a specific example. We proceed from the fact that for the variable \( x_1 \) "transaction time" the value can be LOW, and the variable \( x_2 \) "cost" - the value "GOOD", and so on. In a set that meets several criteria that will be correlated with some values, we can describe the perceptions of persons related to how decisions will be made. These solutions show how satisfactory the alternative would be. We will consider the variable \( S \) "satisfactory" as linguistic. Next, we consider an example of a statement:

\[
d_1: \text{"If } x_1=\text{"LOW" and } x_2=\text{"GOOD"}, \text{then } S=\text{"HIGH"}.\tag{5}
\]

When we consider the general case, then the statement \( d_1 \) is represented as follows:

\[
d_1: \text{"If } x_1 = A_u u x_2 = A_2 u \ldots x_p = A_p, \text{mo } S = B_i.\tag{6}
\]

We introduce the notation of intersection \( (x_1 = A_{u0} u x_2 = A_{u2} u \ldots u x_p = A_{uP}) \) in the form \( x = A_i \).

Operations associated with the intersection of fuzzy sets will correspond to the definition of the minimum by their membership functions: \( \mu_{A_i} (v) = \min (\mu_{A_{u0}} (u_1), \ldots, \mu_{A_{uP}} (u_p)) \). In the specified expression \( V = U_1 \times \ldots \times U_p ; v = (u_1, \ldots, u_p) \mu_{A_{uJ}} (u_j) \) shows what the value of the element's membership \( u_j \) in a fuzzy set \( A_{uj} \) will be. After that, we will write expression (6) as follows:

\[
d_1: \text{"If } x = A_i, \text{mo } S = B_i.\tag{7}
\]

Base sets \( U \) and \( W \) we will denote \( V \) as \( W \). In this case \( A_i \) will be a fuzzy subset of \( W \). When constructing a model \( B_i \) is considered as a fuzzy subset in the unit interval I.

When developing a model, it is necessary to present the rules. In this case, we rely on the operation
of implication. Lukasiewicz’s fuzzy implication will be as follows:

$$\mu_H(w,i) = \min(1, (1 - \mu_A(w) + \mu_B(i)))$$  \hspace{1cm} (8)$$

In this case, $H$ - will be a fuzzy subset on $W \times 1, w \in W, i \in I$.

Similarly to how it was done above, we can make transformations of statements $d_1, ..., d_q$ to sets $H_1, ..., H_q$. The set $D$ is considered as their intersection:

$$D = H_1 \cap \ldots \cap H_q$$

For any of $(w,i) \in W \times I$  $\mu_D(w,i) = \min(\mu_{H_i}(w,i)), i = 1, q$.

To what extent we satisfy the alternative associated with a fuzzy subset of $A$ and $W$, we can deduce using of the compositional inference rule $G = A \circ D$, in it, $G$ is considered as a fuzzy subset over the interval $I$.

Summarizing, we get

$$\mu_O(i) = \max(\min \mu_A(w) \mu_D(w,i)).$$

Alternatives will be generated by generating point estimates. For a fuzzy set $C \in I$ we can find $\alpha$ - level set $(\alpha \in [0,1])$:

$$C_\alpha = \{i | \mu_C(i) \geq \alpha, i \in I\}$$  \hspace{1cm} (9)$$

For any $C_\alpha$ there is a possibility to determine the average number of elements $M(C_\alpha)$:

$$M(C_\alpha) = \sum_{j=1}^{n} i_j / n, i_j \in C_\alpha$$

The set $C_\alpha$ that consists of $n$ elements if $C_\alpha = \{a \leq i \leq b\}, M(C_\alpha) = (a + b)/2; i$

$$C_\alpha = \bigcup_{j=1}^{n} \{a_j \leq i \leq b_j\}, M(C_\alpha) = \frac{\sum_{j=1}^{n} 0.5(a_j + b_j)(b_j - a_j)}{\sum_{j=1}^{n} (b_j - a_j)}, \text{ when } 0 \leq a_j \leq b_j \leq \ldots \leq a_n \leq b_n \leq 1.$$  \hspace{1cm} (10)$$

In this case, we can record a point value over a set $C$ in this way:

$$F(C) = \frac{1}{\alpha_{\text{max}}} \int_{0}^{\alpha_{\text{max}}} M(C_\alpha) d\alpha,$$

$\alpha_{\text{max}}$ is considered as the maximum value over set $C$.

When alternatives are chosen, then it is important for any of them to find satisfaction. The corresponding point estimate is calculated. Which alternative would be considered the best? The whichever has the largest value [13]. A random enumeration is performed based on the index $i_j$. It demonstrates what the numbers will be in the criteria. We introduce a discrete random variable $\tilde{\eta} \sim \tau$ taking values $1, 2, ..., I_1$... The initial distribution is uniform

$$P(\tilde{\eta} = i_j) = \frac{1}{I_1}, \forall i_j = 1, I_1.$$  \hspace{1cm} (11)$$

When changing the distribution (11) using information from the decision maker, we rely on an algorithm that has such steps.

1 step. We implement the random variable $\tilde{\eta} = \alpha$ in a machine way. Then we conduct an experiment using the criterion $\Psi_\alpha(x), \alpha \in \bar{I}, \bar{I}_1$. It is important to follow the procedures. Values are
calculated based on other criteria $\Psi_{il}(x)$ ($i_1 = l, I_1, i_\alpha \neq \alpha$) if $x = x_\alpha^k$. Using the obtained values, we will be able to construct a vector $z^k$:

$$z^k = \begin{pmatrix} \Psi_{1}(x_1^k) \\ \vdots \\ \Psi_{i_l}(x_{i_l}^k) \end{pmatrix}.$$  \hspace{1cm} \text{(12)}

Step 2. It is necessary to involve the person who will make the decision. For such a person, the appropriate question is formulated: "Will we get satisfactory values for all the components in the vector $z^k$?" We will complete the solution of the problem under consideration if there is a satisfactory answer. Otherwise, an analysis is required regarding the possibilities for folding local criteria to the global one [14]. It will look like this:

$$F(x) = \sum_{i=1}^{I_1} P_{il}^* \hat{\Psi}_{il}(x).$$  \hspace{1cm} \text{(13)}

In the specified expression $P_{il}^*$ - will be the steady-state values of probabilities

$$P^k(\bar{i} = i_1) \left( i_1 = \overline{1, I_1} \right), \quad \hat{\Psi}_{il}(x) = \frac{\Psi_{il}(x) - \Psi_{il}^{\min}(x)}{\Psi_{il}^{\max}(x) - \Psi_{il}^{\min}(x)},$$  \hspace{1cm} \text{(14)}

$\Psi_{il}^{\max}, \Psi_{il}^{\min}$ - for local criteria the maximum and minimum values. In this case, they will lie inside in the areas of feasible solutions. How does this happen? Uniform distribution is considered for the first step

$$P^k(\bar{i} = i_1) = \frac{1}{I_1}, \left( i_1 = \overline{1, I_1} \right).$$  \hspace{1cm} \text{(15)}

The preferences of the decision-makers are changing. Therefore, there will be changes in the values of the probabilities. But their total sum will be as follows

$$\sum_{i=1}^{I_1} P^k(\bar{i} = i_1) = 1.$$  \hspace{1cm} \text{(16)}

When the process of learning situations is carried out, then there will be a tendency of the value $P_{il}^*$
of probabilities (figure 2) to certain values that are steady-state [15].

**Figure 2.** Illustration of a diagram of the settings of probabilistic estimates associated with preferences by local criteria.

A visual registration of the person who makes the decision takes place. It will be indicated to them as a function to be optimized by machine experiment. In this case, we rely on a procedure when criterion (16) is applied. When the decision-maker believes that the values \( P^k (\tilde{\eta} = i) \) \((i = 1, I)\) will not be outside the ranges of steady-state values, then he marks the criterion \( \Psi_{\beta}(\chi_{\alpha}) (\beta \in 1, I) \) in which the least satisfactory value will be. We can interpret his answer in the \( k \)-th step in the form of signed estimates for the values

\[
\Delta_{i} F(x_{a}) = \text{sign} \left( \frac{\partial F(\Psi)}{\partial \Psi_{i1}} \right) = \begin{cases} 1, i = \beta, \\ -1, i \neq \beta. \end{cases}
\]

(17)

Step 3. The settings process will be carried out according to the values of the probabilities \( P^{k+1} (\tilde{\eta} = 1) \). They correlate with how the generation of random discrete variables \( \tilde{\eta} \) for the \((k+1)\)-th step will proceed:

\[
P^{k+1} (\tilde{\eta} = i) = \frac{P^k (\tilde{\eta} = i) + \varepsilon^{k+1} \varepsilon \left( \Delta_{i} F(x_{a}) \right)}{1 + \varepsilon^{k+1}} \quad (i = 1, I),
\]

(18)

In this case \( \varepsilon^{k+1} \) - is considered in the form of a step size correlated with an iterative procedure

\[
(a) = \begin{cases} 1, & \text{if } a \geq 0, \\ 0, & \text{if } a \leq 0. \end{cases}
\]

(19)

in the numerator we mean an iterative gradient procedure. The expression written in the denominator shows how the condition is satisfied

\[
\sum_{\tilde{\eta}} P^{k+1} (\tilde{\eta} = i) = 1.
\]

(20)

When the step size \( \varepsilon^{k+1} \) is chosen, then we are based on heuristic considerations. When a person...
who will make a decision on two adjacent steps \((k-1)\) and \(k\) will make a choice according to the same criterion \(\beta\), then we can assume that there will be an increase in its significance according to experimental research. In such cases, the \( P^{k+1}(\tilde{\eta} = \beta) \) will be growth. We can ensure this by increasing in steps \(\varepsilon^{k+1}\). In cases where the preferences are not the same, it is required to achieve opposition relative \(\varepsilon^{k+1}\). How can this be explained? Then the persons associated with how decisions are made are not characterized by a stable opinion. There will be an adjustment of opinions. And this leads to the fact that there is no need to carry out a noticeable correction with respect to the distribution over a discrete random variable \(\tilde{\eta}\). That is, when a choice is made according to the step size, we rely on the rule

\[
\varepsilon^{k+1} = \varepsilon^k \exp\left[\gamma \Delta^k \beta \Delta^{k-1} \beta \right]
\]

In the expression \(\gamma\) - considered as a constant factor. After setting the probabilities, return to the first stage.

4. Conclusion
Distributed database must undergo a process of structural synthesis. This is necessary to improve the characteristics of the CIS. Although optimization procedures are performed, the values of the objective function are calculated. For this, a simulation model is involved. It provides opportunities to formalize the dependence of the query processing time on alternative variables. The use of alternative variables is necessary in tasks related to the synthesis of the number of parallel processing and priorities for fulfilling certain processing conditions. When the process of selection of the final options for the design is carried out according to a variety of promising options, then the involvement of a multi-criteria assessment is required. Formalization of the specified requirements in a fuzzy way allows improving its efficiency. This introduces linguistic variables. The algorithm uses a randomized scheme associated with the search for a compromise solution.

References
[1] Stankovic J A 2014 Research directions for the internet of things IEEE Internet of Things Journal 1 3-9
[2] Rios L M and Sahinidis N V 2013 Derivative-free optimization: a review of algorithms and comparison of software implementations Journal of Global Optimization 54 1247-93
[3] Lutakamale A S, Kajjage S 2017 Wildfire Monitoring and Detection System Using Wireless Sensor Network: A Case Study of Tanzania Wireless Sensor Network 9 274-89
[4] Minerva R, Biru A and Rotondi D 2015 Towards a definition of the Internet of Things (IoT) IEEE Internet Initiative Torino, Italy Retrieved from: https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Re vision1_27MAY15.pdf
[5] 2016 Broadband Internet Technical Advisory Group Internet of Things (IoT) Security and Privacy Recommendations Retrieved from: https://www.bitag.org/documents/BITAG_Report_-_Internet_of_Things_(IoT)_Security_and_Privacy_Recommendations.pdf
[6] Odu G O and Charles-Owaba O E 2013 Review of Multi-criteria Optimization Methods Theory and Applications 3 1-14
[7] Shah A and Gahramani Z 2015 Parallel predictive entropy search for batch global optimization of expensive objective functions Advances in Neural Information Processing Systems 3330-8
[8] Neittaanmki P, Repin S and Tuovinen T (Eds.) 2016 Mathematical Modeling and Optimization of Complex Structures; Series: Computational Methods in Applied Sciences Springer International Publishing AG, Switzerland
[9] Morais H, Kádár P, Faria P, Vale Z A and Khodr H M 2010 Optimal Scheduling of a Renewable Micro-Grid in an Isolated Load Area Using Mixed-Integer Linear Programming Elsevier
Editorial System(tm) for Renewable Energy Magazine 35(1) 151-6

[10] Orlova D E 2018 Stability of solutions in ensuring the functioning of organizational and technical systems Modeling, Optimization and Information Technologies 6(1) 325-36

[11] Talluri S, Kim M K and Schoenherr T 2013 The relationship between operating efficiency and service quality: are they compatible? Int. J Prod Res 51 548-2567

[12] Groefsema H, Beest N R T P 2015 Design-time compliance of service compositions in dynamic service environments Int. Conf. on Service Oriented Computing & Applications 108-15

[13] Lvovich I Ya, Lvovich Ya E, Preobrazhenskiy A P, Preobrazhenskiy Yu P, Choporov O N 2019 Modelling of information systems with increased efficiency with application of optimization-expert evaluation Journal of Physics: Conf. Ser. Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations; Polytechnical Institute of Siberian Federal University. Bristol, United Kingdom 33079

[14] Yao Y and Chen J 2010 Global optimization of a central air-conditioning system using decomposition-coordination method Energy and Buildings 5 570-83

[15] Lvovich I, Preobrazhenskiy A, Preobrazhenskiy Y, Lvovich Y, Choporov O 2019 Managing developing internet of things systems based on models and algorithms of multi-alternative aggregation Int. Seminar on Electron Devices Design and Production, SED-2019 - Proceedings 8798413