Computational approach to synthesis of the multiversion structure of distributed information decision-making support system

I V Kovalev1,2,3,4,8, N A Testoyedov4,6, A A Voroshilova1,5, D I Kovalev2,5 and D V Borovinskii7

1 Siberian Federal University, Krasnoyarsk, Russia
2 Krasnoyarsk State Agrarian University, Krasnoyarsk, Russia
3 China Aviation Industry General Aircraft Zhejiang Institute Co., Ltd, China
4 Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia
5 Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations, Krasnoyarsk, Russia
6 JSC “Academician M F Reshetnev Information satellite systems”, Zheleznogorsk, Krasnoyarsk region, Russia
7 FSBEE HE Siberian Fire and Rescue Academy EMERCOM of Russia, Zheleznogorsk, Krasnoyarsk region, Russia

8 E-mail: kovalev.fsu@mail.ru

Abstract. The computational approach to synthesis of the multiversion structure of distributed information decision-making support system is presented. A formal model of the local information system is given. This system is intended to ensure the functioning of a complex control system based on the multiversion approach and consisting of a set of multiversion objects. The problem of distribution of objects by local information subsystems has been solved. For a set of valid queries in a distributed system, the answers for the decision maker are formed sequentially without repeating queries. To take into account certain requirements regarding the structure of a distributed system, it is necessary to formulate these requirements in formalized constraints and introduce them into the mathematical description of the problem. Note that the effectiveness of the targeted use of the system depends both on the results of synthesis (structure and parameters of the system) and on the correct organization of the subsystem for monitoring its technical condition during operation.

1. Introduction
At present, in the study of complex systems, structural-algebraic methods are widely introduced. In [1-3], a group of methods based on the category-functor approach was considered. This is explained by the fact that the very concept of "category" naturally combines the properties of systematicity and integration of structures. This approach is reflected in the work [4], in which the general system principles of category-functor analysis are considered. This article discusses the application of this approach in the synthesis of multiversion structures of distributed information systems for decision support [5-8].
Let be \( S_j = (T_j, D_j, M_j, \delta_j), j = 1, \ldots, v \) - local information systems (local information systems - LIS),
designed to ensure the functioning of a complex control system (control system - CS), made on the basis of a multiversion approach and consisting of a set \( K = \{1, \ldots, \phi\} \) multiversion objects.

Information from each local information system goes to the information collection center for further transmission to the control center. The multiversion of the interaction of system objects is ensured by the fact that one local information system can serve several objects of the control system (systems of the upper level of the hierarchy).

Distributed information system (Distributed information system – DIS) \( S = (T, D, M, \mu) \) defined through local components:

1) \( T = \bigcup_j T_j; R_j = R \cap (T_j \times T_j); K_j = K \cap (T_0 \times T_0); \)
2) \( D = \bigcup_j D_j; \)
3) \( (K \supset \bigcup_j K_j) \cap (K_j \mu = K \cap (M_j \times M_j)); \)
4) \( \forall m \in M \delta(m) = \{d : d \in D \land m \subseteq t(d)\}. \)

The distributed information system acts as a center for collecting and processing information, i.e. processing responses to requests from governing bodies.

2. Formulation of the problem

To formalize the problem, we introduce the notation: \( c_d \) - the cost of collecting information about the \( l \)-th object of the \( j \)-th local information system; \( b_d \) – the cost of transferring a unit of information about the \( l \)-th object to the center from the \( j \)-th information local system; \( x_j \) – a boolean variable, equal to 1 if the \( l \)-th object is served by the \( j \)-th local information system, and equal to 0 otherwise.

The description of the state of each object is presented in the form of its own information model: \( S_j = (T_j^l, D_j^l, M_j^l, \delta_j^l), l \in K, j \in \{1, \ldots, v\}, \) where \( T_j^l \) - descriptor set thesaurus \( T_{0j}^l \), describing the state of the \( l \)-th object; \( D_j^l \) - a collection of possible documents that governing bodies need to make a decision; \( M_j^l \) - many valid requests from the governing bodies; \( \delta_j^l : M_j^l \rightarrow 2^{D_j^l} \) - display mapping each question to multiple documents.

Index \( j \) indicates that the information model of the \( l \)-th object is formed in the \( j \)-th local information system.

If a local system serves several objects, then it is formally a distributed system of a “point” multiversion type.

Multiversion information about objects is presented by independent information models and at the same time it is concentrated in one place, in one “point”. Thus, information models of objects are subsystems (or multiversions) of the local system serving these objects. This makes it possible to increase the information reliability of the interaction of structural elements of systems and to formulate responses to queries in the form

\[
\delta_j (m) = \{d : d \in D_j \land m \subseteq t(d)\}, \quad (1)
\]

where \( m \in M_j \).

The correctness of the answer is guaranteed by the multiversion properties of the distributed system presented above.

For complex queries that include several descriptors, expression (1) has the form
\[ \delta_j(m) = \bigcap_{i=1}^{k} \bigcup_{i=1}^{p} \delta_j^i(m) = \bigcap_{i=1}^{k} \bigcup_{i=1}^{p} \{ d : d \in D_j^i \land m_j \not< t(d) \}, \]  

where \( m = \{m_1, \ldots, m_k\}, m_i = \{m_i\} \); \( p \) – the number of objects served by the local information system.

The volume of information transmitted to a request \( m \) from the \( j \)-th local information system is equal \( \alpha_j = F(\delta_j(m)) \), to where \( F \) is the operator of converting information to the form intended for transmission to communication channels.

3. Solution method

Let us formulate the problem of distributing objects over local information subsystems with a set of admissible requests in a distributed system \( M = \{m_1, \ldots, m_r\} \), to which responses are formed sequentially without repeating requests. For convenience, let us assume that each request is described by one descriptor.

It is required to find

\[
\min \sum_{i=1}^{k} \sum_{i=1}^{p} c_{ji} x_{ji} + \sum_{i=1}^{k} \sum_{i=1}^{p} b_{ji} F(\delta_j^i(m_j)) x_{ji}
\]

with restrictions

\[
\sum_{j=1}^{J} x_{ji} = 1, \forall i \in K; \tag{4}
\]

\[
\sum_{j=1}^{J} x_{ji} \geq (\leq)N_j, \forall j \in [1, v]. \tag{5}
\]

Constraint (4) requires servicing each object with only one information system. Condition (5) limits the number of objects to be serviced by local subsystems, or, on the contrary, requires that they be at least a given number.

The considered by us multiversion approach to the synthesis of the structure of a distributed information system for decision support makes it possible to complicate the requirements for a distributed system.

We will demand that the governing bodies receive answers to requests even if any local information system ceases to function. In other words, there is a need for duplication (or multiversion representation) of information about objects.

The statement of the problem, taking into account the multiversion, changes only in the part concerning the change in constraint (4). It takes the form

\[
\sum_{j=1}^{J} x_{ji} = 2, \forall i \in K. \tag{6}
\]

Let us modify the problem statement again. We will assume that after assigning objects to local information systems in problem (3) - (5), the objects are not redistributed. However, the control bodies need to have information about the previous states of objects even after the termination of the functioning of any of the local subsystems. This means that there is a duplication of information models about objects constantly in the process of their functioning and the information of local information systems is redistributed between them.

For clarity of presentation, let us simplify the situation, assuming that the redistribution of information is carried out once. Although in a real situation information about the state of objects is transmitted as it arrives over the entire time interval of the operation of local information systems, this restriction does not reduce the generality of reasoning, since the complete formulation of the problem will simply require additional summation over discrete times.
Let us introduce the following notation: $a_{jl}$ – the cost of transferring a unit of information from the j-th local information system to the s-th; $y_{jl}$ – boolean variable, equal to 1 if information about the l-th object is redistributed from the j-th local information system to the s-th.

Finally, the problem of synthesizing the structure of a distributed information system is formulated as follows: it is necessary to find

$$\min_{\{x_{jl}, y_{jl}\}} \sum_j \sum_l c_{jl} x_{jl} + \sum_j \sum_l \sum_s b_{jl} F(\delta_j(m)) x_{jl} + \sum_j \sum_l \sum_{s} a_{jl} [F(T_j^l, D_j^l, M_j^l, \delta_j^l) x_{jl}] y_{jl}$$

(7)

with restrictions

$$\sum_j x_{jl} = 1, \forall l \in K;$$

(8)

$$\sum_j x_{jl} \geq (\leq) N_j, \forall j \in \{1,...,v\}$$

(9)

$$\sum_j y_{jl} = 1, \forall l \in K_j = \{l : x_{jl} = 1\}, \forall j \in \{1,...,v\}.$$

(10)

Constraint (10) requires the redistribution of information about the l-th object from the j-th local information system.

4. Discussion of results

Based on problem (7) - (10), a number of problems can be formulated that take into account certain requirements regarding the structure of a distributed system. To do this, it is necessary to formulate the requirements in the form of restrictions and enter them into the description of the problem.

Let us consider what is new in the formulation of problems of synthesis of distributed systems by the use of a system model on a thesaurus with weights. The main advantage of this model, as already noted, is the ability to take into account the distribution of information over the descriptors that describe the document, and thus, in fact, the ability to assess the uncertainty in describing the responses to requests from the management bodies.

So, let documents in local information systems be represented in the form

$$t(d_j^l) = \{<t_{1j}, w_1>, <t_{2j}, w_2>, ..., <t_{kj}, w_k>\},$$

$$d_j^l \in D_j^l, S_j^l = (T_j^l, D_j^l, M_j^l, \delta_j^l).$$

In real situations, when describing the state of objects, it is natural to expect that different local information systems, provided with different technical means of measuring and monitoring the state of objects, will represent descriptions of the same objects that do not coincide with each other. Then upon request $\bar{m}$ controls about the state of a certain object $l$ local information systems provide answers with similarity accuracy $\rho$:

$$\delta_j^l(\bar{m}, \rho) = \{(d_j^l, \alpha_j^l) : d_j^l \in D_j \land \bar{m} \preceq \rho t(d_j^l) \land \alpha_j^l = \mu(\bar{m}, t(d_j^l)), \forall l \in K, j = 1, v.\}

Obviously, the indicators $\alpha_j^l$ can be used to refine the structure of a distributed system, that is, to select a structure that would provide answers with the maximum measure of correlation or with a measure of correlation no less than a given one. The last requirement is equivalent to the condition for transmitting information with a content that meets the request, at least as specified. Formalization of this condition in relation to the formulation of problem (7) - (10) consists in the introduction of additional constraints

$$\alpha_j^l = \mu(\bar{m}, t(d_j^l)) \geq \alpha_j^l, \forall l \in K, \forall j : x_{jl} = 1.$$

(11)
In practical applications, other requirements for the accuracy of the received answers are also possible. It is only necessary to always clearly understand the expediency of using certain models and the final effect of their implementation.

5. Conclusion
In conclusion, we note that the effectiveness of the targeted use of a system for any purpose depends not only on how successfully the structure and parameters of the system were selected during the synthesis, but also on how rationally the system for monitoring its technical condition is organized during operation. For the proper organization of control, already at the stage of system development, the fulfillment of the testability requirements for the developed system should be provided. In [10] it is determined that testability is a property of a product (system, object), which characterizes its adaptability to control by specified means.

Ensuring the controllability of the system requires the solution of rather complex problems, such as the construction of models of the system as an object of control, the choice of controlled features and the substantiation of the composition of control points in the system, optimization of control procedures, etc. At the same time, to solve these problems, the mathematical concepts presented in this article are used and designs, developed and supplemented, if necessary, with new elements in relation to the specific task under consideration.

References
[1] Dmitriev A K and Maltsev P A 1988 Fundamentals of the theory of design and control of complex systems (Sankt Petersburg Energoatomizdat)
[2] Mazur Z 1984 Information Processing and Management 20(4) 499-505
[3] Zenutkin N et al. 2021 Modern Innovations, Systems and Technologies 1(1) 10-22
[4] Leskin A A 1986 Algebraic models of flexible production systems (St. Petersburg: Nauka)
[5] Losev V V et al. 2020 IOP Conf. Ser.: Earth Environ. Sci. 548 032042
[6] Losev V V et al. 2020 J. Phys.: Conf. Ser. 1679 052088
[7] Kovalev I V et al. 2020 Information and Software Technology 120 106245
[8] Saramud M V et al. 2019 Devices and Systems: Management, Control, Diagnostics 3 19-25