The study of the complex information systems structure on the coefficient of structural uncertainty basis

V V Alekseev¹, A V Yakovlev¹, F E Aminova² and M Ju Sergeev³

¹Tambov State Technical University, 106 Sovetskaya Street, Tambov, 392000, Russian Federation
²Moscow Aviation Institute (National Research University), Volokolamsk highway, 4, Moscow, 125993, Russian Federation
³Voronezh State Technical University, 14, Moscow ave., Voronezh, 394026, Russian Federation

E-mail: vvalex1961@mail.ru

Abstract. An approach to the analysis of the topological structure of a system on the basis of the coefficient of structural uncertainty, which allows to assess an ability to modify the structure and compare different topologies, is considered in this paper. The coefficient is defined by the number of transit and end vertices of a graph, which describes the system under consideration. The task is to find a system structure with the greatest value of the coefficient of structural uncertainty. The limits regarding information loss during destruction of one system element with a fixed number of elements is taken into account. The results of research on inter-element communication recombination and a number of elements within a level of the hierarchy that both influence the structural uncertainty parameter are provided. A sensitivity analysis of the coefficient of structural uncertainty related to a transit vertex topology is carried out.

1. Introduction

The development and increasing complexity of systems of information processing, management, or communication systems are following the upward trend [1]. Hence, a great deal of Russian and foreign research projects is dedicated to formal models of complex system structures. However, it would definitely be too early to state that the solution has been found since these are multi-faceted and compound systems. The systems have common features such as a great number of elements, plethora of executing functions, connections and interactions between them, problem of selecting a controlling influence algorithm, and significant amount of processing information. The most prominent features are the following:

- complexity and structure type (emerged due to of the elements): hierarchical, regular, multilayer (rarely are there weak influences capable of increasing the complexity);
- system dynamics, complexity of its behavior, prediction ability;
- presence of patterns [2]: interactions between parts and the entity (integrity), hierarchical order of systems (communication, hierarchy), functionality and development of systems (self-
organization, history); system implementation (equifinality, the law of requisite variety, potential efficiency)

- weakness or the lack of the structure.

An analysis of different aspects that influence the formation or choice of a complex technical system structure is provided in references [2], [3], [4]. They can be classified into two groups:

- the first one is related to the system hierarchy: number of levels of the hierarchy, types of the interaction, degree of the centralization, aspects of system division into subsystems;
- the second one is related to the efficiency of its functionality: efficiency, reliability, liveliness, system performance, adaptability.

2. Problem definition of a system structure assessment

It is obvious that during the process of assessment of topological structure complexity it’s vital that the ratio between terminal and transit nodes be considered [5], [6]. We introduce a definition of a transit node in terms of the problem. Considering the given limits, an unconventional approach to complex information system structure research is required. A set of $U$ nodes of a graph $G$ can be divided into three groups by their aspects of connectedness:

\[ U = \{\alpha_1, \alpha_2, ..., \alpha_4\} \cap \{\beta_1, \beta_2, ..., \beta_4\} \cap \{\gamma_1, \gamma_2, ..., \gamma_4\} \] (1)

where \{\alpha_1, \alpha_2, ..., \alpha_4\} is a subset of transit nodes; \{\beta_1, \beta_2, ..., \beta_4\} is a subset of terminal (input) nodes (information consumers); \{\gamma_1, \gamma_2, ..., \gamma_4\} is a subset of root output nodes (information sources). The number of input nodes shouldn’t be greater than 1, and the number of branches of transit nodes shouldn’t be less than 2.

A problem of the analysis of complex information systems on the basis of the coefficient of structural uncertainty may be as follows:

Consider $L$ options suggested for a complex system structure described by a graph $G$.

For each graph $G$: $\Lambda, M, N, w_{1\alpha}, w_{1\beta}$ where: $\Lambda$ is a number of transit nodes; $M$ is a number of terminal nodes (information consumers); $N$ is a number of root nodes (information resources); $w_{1\alpha}$ is a reachability matrix of transit nodes for a system option $l$; $w_{1\beta}$ is a reachability matrix of end nodes for a system option $l$.

Find a structure option for the complex information system where the coefficient of structural uncertainty $\Psi_l$ complies with $\Psi_l \xrightarrow{G_l \to} \max$.

The main feature of the developed approach to complex information system structure research is a method that determines a parameter reflecting the distinction between multitudes of control system elements, i.e. system nodes obtaining a specific structure called the coefficient of structural uncertainty.

The concept is based on the following axiom: during the process where a number of transit nodes reaches the total number of nodes in an information system, in which its nodes have approximately the same weight, the coefficient of structural uncertainty tends to 0; meanwhile, the rate of information losses tends to the number of the same rate but of the below level. The following formal equation is used to compute the coefficient of structural uncertainty [6]:

\[ \Psi = 1 - \frac{\sum_{i=1}^{\Lambda} w_{\alpha_i} \cdot \alpha_i}{\left(\sum_{i=1}^{\Lambda} w_{\alpha_i} \cdot \alpha_i + \sum_{j=1}^{M} w_{\beta_j} \cdot \beta_j + \sum_{k=1}^{N} \gamma_k\right)} \] (2)
where \( A \) is a number of transit nodes; \( M \) is a number or terminal nodes (information consumers); \( N \) is a number or foot nodes (information resources); \( w_\alpha \) is a coefficient of reachability (the weight of a transit node \( i \)):

\[
w_\alpha = \sum_{i=1}^{n} p_{\alpha i},
\]

where \( p_{\alpha i} \) is a number of successors of a transit node \( i \), \( r \) is a number of connections of a transit node \( i \) for all its successors; \( \beta_i = \frac{1}{(M + N)} \) is an end node coefficient.

Note that the coefficient of structural uncertainty for complex information systems with a unique structure may vary due to different numbers of the abovementioned node types. It also may not reflect structural features of the system, i.e. a number of nodes in a level, node distribution of a level \( z_s \) in respect to the levels \( z_{s-1} \) and \( z_{s+1} \). Take into account the fact that structural functionality reliability research is the prime problem of such research, i.e. vulnerability detection within a system topology and its influence on the system’s integrity, thus an additional parameter is to be defined. The parameter describes the homogeneity of node distribution of lower level \( z_{s+1} \) with respect to upper layer \( z_s \). The parameter influences the nodes destruction probability (subtrees) of layer \( z_{s+1} \).

Thus, in terms of resolving the problems of complex information system structure research we consider the given system containing the maximum reliability (with regard to the topology) if the coefficient of structural uncertainty \( \Psi \) is close to zero and doesn’t change during the nodes transformation.

The coefficient of structural uncertainty \( \Psi \) resembles a hash function since it:

- requires few calculations using basic transformations;
- allows to measure complexity of topologies of different structures;
- implies high complexity; impossible to estimate features of the system;
- might be equal across all structures with different topologies.

Ultimately, complete functionality reliability research entails vulnerability verification which also implies some modelling. The modelling should be conducted on the basis of criteria selection that will not include the conditions of complete indeterminacy and does suppose that the occurrence of undesirable influence may be predicted. By doing so, it will provide an opportunity to choose the one that brings the best results.

Let \( \Theta_1, \Theta_2, \Theta_p \) be different groups of unwanted influence on objects of complex information system control \( O_1, O_2, O_j \). A group \( \Theta_i \) consists of the following elements \( D_{t_1}, D_{t_2}, D_{t_3} \) such that \( \Theta_i = \{D_{t_1}, D_{t_2}, ..., D_{t_3}\} \). The power of effect of every element of the group \( D_{t_3} \) influencing object \( O_j \) by \( \Theta_i \) can be evaluated with:

\[
f_{j}(D_{t_3}, O_j) = U(i, k, j) \geq 0,
\]

where \( U \) is a nonempty set of nodes.

External influence \( \Theta_i \) within a discrete period of time where \( i = 1, 2, 3, ..., m \), is evaluated with the following equation:

\[
Y_i(t) = U(t) - U(t - 1),
\]

where \( t > 0 \).

If \( t > 0 \), then the result of the external influence for the \( i \)-th vertex of the \( G \) graph can be evaluated with the following equation:

\[
S_i(t+1) = S_i(t) + \sum_{t=1}^{k} S_i Y_i(t),
\]
where $S_i$ is an amount of lost information after losing the $i$-th node, also take into account that $k$ is a number of nodes adjacent to the $i$-th node, that are the beginning of $X$ arcs.

If there exists an external influence, then the following two events are inherent to elements and system connections:

- Bifurcation as an unpredictably divided element;
- Recombination of connections between elements as a process opposed to destruction.

3. **Research on inter-element communication influence on the coefficient of structural uncertainty**

The following table illustrates the connection recombination process in a hierarchical system. This process turns the system into a network one [6], [7]. The respective parameter assessment in which the connection direction was taken into account was carried out. Yet such results are not listed in the table 1.

Then we construct an efficiency matrix in which information losses of the complex information system are calculated. This matrix depends on the amount of unwanted influence ($D_i$) and a possibility of node destruction of the network. Frequencies, which define the propagation and correspond to the amount of unwanted influence, and decision-making regarding optimization criteria should be taken into account.

**Table 1.** Inter-element communication recombination influence on the coefficient of structural uncertainty $\Psi$.

| № | Structure type | $\Psi$ | № | Structure type | $\Psi$ |
|---|----------------|-------|---|----------------|-------|
| 1 | ![Structure](image1) | 0.93  | 4 | ![Structure](image2) | 0.805 |
| 2 | ![Structure](image3) | 0.88  | 5 | ![Structure](image4) | 0.696 |
Information losses that were dependent on the number of influences were experimentally obtained. Their weight wasn’t considered. The general parameter can be evaluated by the following recurrence relation:

\[ S_n = \frac{n^* S_{n-1} - S_n}{n-1}; \quad S_1 = C. \]

System information includes:

- a parameter that depends on the number of system elements;
- a parameter that depends on the way elements interact with each other.

Research on the coefficient of structural uncertainty for calculating such dependent parameters was carried out. Two types of hierarchical structures were studied [6], [7], [8]: B-trees (binary branching) and T-trees (ternary branching) by consistently increasing the level in the hierarchy and calculating the coefficient of structural uncertainty \( \Psi \).

The following table 2 illustrates the process of how this coefficient changes.

**Table 2.** The influence of the number of elements in a hierarchical level on the coefficient of structural uncertainty \( \Psi \).

| No | \( B \)-trees | \( \Psi_B \) | \( T \)-trees | \( \Psi_T \) |
|----|----------------|------------|---------------|------------|
| 1  |                | 0.733      | 1             | 0.709      |
| 2  |                |            | 0.809         |            |
| 6  |                | 0.489      |               |            |
The results were used to plot the graph. The numbers specify how the coefficient of structural uncertainty $\Psi$ and information losses depend on the number of elements in a level of the hierarchy.

Figure 1. Dependence of the coefficient of structural uncertainty $\Psi$ and information losses on the number of elements in a level of the hierarchy.

4. Conclusion
The following provisions can be deduced after doing the research:

- the coefficient of structural uncertainty $\Psi$ resembles a hash function: requires few calculations using basic transformations; allows to measure complexity of topologies of different
structures; implies high complexity; impossible to estimate features of the system; might be equal across all structures with different topologies;

- the coefficient of structural uncertainty for complex information systems with a unique structure may vary due to different numbers of the node types (transit, terminal, root). It also may not reflect structural features of the system;
- the coefficient of structural uncertainty allows to carry out research on the structural functionality reliability parameter, i.e. detecting topology vulnerabilities of a system and defining the extent to which it influences system integrity;
- the coefficient of structural uncertainty is sensitive to the number of system-defining elements and the number of hierarchy levels; hence, this results in a much higher level of comprehensiveness and a much more prominent effect; yet, there’s a trend towards decreasing differences in the parameters of the coefficient of structural uncertainty if the number of hierarchy levels increases.

References

[1] Voronin A A and Mishin S P 2003 *Optimal hierarchical structures* (M.: IoCS RAS)
[2] Baranov V V and Tsvirkun A D 2018 Development control: structural analysis, problems, stability *Autom. Remote Control* **79**(10) 1780–96
[3] Gubko M V 2006 *Mathematical models for optimizing hierarchical structures* (M.: Lenand)
[4] Simankov V S and Lutsenko E V 1999 *Adaptive management of complex systems on the basis of the theory of pattern recognition* (Krasnodar: KubSTU)
[5] Burkov V N, Korgin N A, Novikov D A and Gubko M V 2015 *Introduction to theory of control in organizations* (New York: CRC Press)
[6] Gubko M V and Mishin S P 2008 *Optimal Hierarchies in Firms: A Theoretical Model* (Seoul: International Federation of Automatic Control) pp 2962-7
[7] Harary F and Palmer E 1973 *Graphical Enumeration* (New York: Academic Press)
[8] Schwenk A J 1973 *Almost All Trees are Cospectral* *New Directions in the Theory of Graphs / F Harary* (New York: Academic Press) pp 275–307