Mathematical model of the architecture of a distributed information-measuring system based on cloud and fog technologies

S P Vorobyev
The Platov South-Russia State Polytechnic University (NPI), 132 Prosvesheniya ul. Novocherkassk 346428 Russia
E-mail: vsp1999@yandex.ru

Abstract: This article describes the formulation of a mathematical model for optimizing the structure of a distributed information-measuring system based on a multi-level topology. The paradigm of cloud, fog and boundary computing is used, which allows building a modern network infrastructure within the framework of the industrial IoT concept. The most typical example of the use of this paradigm may be intelligent power supply networks and distributed systems for monitoring the transportation of gas or oil products. To solve the problem, it is proposed to use an approach based on a genetic algorithm.

1. Introduction
Currently, the urgent need for optimal design of distributed information and measuring systems of multifunctional measuring complexes is caused by the number of reasons. They include the dramatic complication of technological processes in production, the need to collect technological and control parameters within a fairly extensive geographically distributed complex, monitoring parameters under conditions of uncertainty, under the influence of destabilizing factors, requirements to high metrological level control. Another reason is the fact that in production and in scientific research information-measuring systems are used that do not fully meet all the necessary technical characteristics: efficiency, a wide range of parameters studied, the ability to control the set of determined parameters, the required metrological level of measurement results. Quite a large number of existing information-measuring systems require the solution of such problems as increasing resistance to the effects of destabilizing factors, functioning under conditions of uncertainty, the possibility of choosing a control method and changing the structure of an information-measuring system as part of a technological process.

Typical examples of a geographically distributed multifunctional measuring complex can serve as a set of electricity metering devices - a set of devices for measuring and metering electricity (current and voltage transformers, electricity meters, pulse sensors, adders and communication lines) as part of the transformation of the power system into a decentralized structure or assessment complex controlled indicators in a geographically distributed system of monitoring the transportation of gas or oil products through. Such complexes, as a rule, provide: telemetry and broadcasting of technological parameters; state telecontrol, telecontrol of crane nodes, diagnostics; archiving and documenting information and operator actions.
Modern trends in the design and development of distributed information systems of multifunctional measuring complexes assume, taking into account the proposed advantages of IoT technologies, their construction based on architectural solutions of cloud-based, fog and boundary computing.

Cloud computing is the concept of building a distributed system with the provision of network access to a scalable and flexible set of common physical or virtual resources. Fog computing is a structure that is located in the common network architecture between the end devices (sensors and measuring system devices) and data centers, which allows to improve service quality, delay time and reduce traffic volume through public networks. Edge computing is a further development of technologies in which the computing infrastructure is even more close to the end client, and most of the operations are performed in the boundary network of a particular client. If in fog information systems, calculations are performed in a node near the place of data collection, then the boundary ones allow data to be processed in the same place where they are received, which allows for instant response, since in this case, a programmable logic controller is used, which can be part of a production or process control system. It controls the end devices and has the ability to instantly process the received data.

Designing the optimal structure of an intricate interconnected complex of software and technical solutions for cloud, fog and boundary computing can be implemented by means of a multi-level representation of the topology of a computer network [1, 2].

2. Methodology

The formulation of the task of optimizing the multilevel topology of a computer network of a distributed information-measuring system of a multifunctional measuring complex from the point of view of the paradigm of cloud, fog and boundary computing levels can also include the task of distributing servers, services and functional subsystems, as well as corresponding information resources across the “cloud”, “fog and boundary” systems.

In this case, to describe the distributed information system of the multifunctional measuring complex $S$ from the point of view of the multilevel approach, the pair $(O^{(l)}, T^{(l)})$, is also introduced, consisting of the set $O^{(l)}$, which contains all the necessary objects $S$, and $T^{(l)}$ is the family of subsets of the set $O$, which define the topological structure systems $S$ at the $l$-th level $l = \{el, fl, cl\}$, respectively, for the levels of boundary, fog and cloud computing. Additionally, the set of functional subsystems $FSS = \{FSS_i, i=1,m\}$; the set of information resources of the functional subsystem $FSS_iIM = \{IM_j, j=1,kir, \}, i=1,m$; the set of servers of the distributed information-measuring system $SR = \{SR_j, j=1,ks\}$. The $FTR(O^{(l)}, T^{(l)})$ function sets the volume of transit traffic at the $l$-level.

At the level of boundary computing, it is necessary to determine such a set of $T_{i_{\text{opt}}}^{(el)}$, which will include the location of servers and services, functional subsystems and their information resources in an optimal way from the point of view of the selected criterion

$$g^{(el)}(T_{i_{\text{opt}}}^{(el)}, SR^{(el)}, FSS^{(el)}, IM^{(el)}) \rightarrow \min \max$$

subject to restrictions $h_j^{(el)}(T_{i_{\text{opt}}}^{(el)}, SR^{(el)}, FSS^{(el)}, IM^{(el)}) \leq H_j^{(el)}, j = 1, m^{(el)}$, where $m^{(el)}$ the number of restrictions imposed on the topological structure of the computer network at the level of boundary computing of the selected interaction model.

At the level of fog computing, in a similar way, it is required to define a subset $T_{i_{\text{opt}}}^{(fl)}$ from the $T^{(fl)}$ family that will describe the entry of elements of active network equipment $X^{(fl)}$ into a distributed structure based on switches, as well as the necessary servers and services and information resources of functional subsystems with

$$g^{(fl)}(T_{i_{\text{opt}}}^{(fl)}, SR^{(fl)}, FSS^{(fl)}, IM^{(fl)}) \rightarrow \min \max$$

in compliance with functional limitations as well as network standard limitations $h_j^{(fl)}(T_{i_{\text{opt}}}^{(fl)}, SR^{(fl)}, FSS^{(fl)}, IM^{(fl)}) \leq H_j^{(fl)}, j = 1, m^{(fl)}$.
At the cloud computing level, the set $X^{(cl)}$ includes elements of the active network equipment, as well as the necessary servers and services and information resources of the functional subsystems, and it is necessary to define a subset $T^{(cl)}_{opt}$ with the optimal criterion value

$$g^{(cl)}(T^{(cl)}_{opt}, SR^{(cl)}, FSS^{(cl)}, IM^{(cl)}_{i}) \to minmax$$

and compliance with the relevant functional limitations, including the quality of service

$$h^{(cl)}_{j}(T^{(cl)}_{opt}, SR^{(cl)}, FSS^{(cl)}, IM^{(cl)}_{i}) \leq H^{(cl)}_{j}, j = 1, m^{(cl)}$$

Detailed mathematical model presumes the set of workstations $WS = \{WS_i, i = 1, n_{ws}\}$, where $n_{ws}$ is the number of workstations; the set of elements of active network equipment (switches, routers, etc.) $SW = \{SW_i, i = 1, n_{sw}\}$, where $n_{sw}$ is the number of elements of active network equipment; the set of servers $SR = \{SR_i, i = 1, n_{sr}\}$, $n_{sr}$ is the number of servers; the set of storage systems $SS = \{SS_i, i = 1, n_{ss}\}$, $n_{ss}$ is the number of storage systems. As a result, the set of network objects $O = WS \cup SR \cup SW \cup SS$.

To build the model, the following variables are introduced:

- $X^{el}_{O_i} = \begin{cases} 1, & \text{if } i\text{-element of the set } O \text{ belongs to the level of boundary computing} T^{el} \\ 0, & \text{otherwise} \end{cases}$

- $X^{fI}_{O_i} = \begin{cases} 1, & \text{if } i\text{-element of the set } O \text{ belongs to the level of fog computing} T^{fI} \\ 0, & \text{otherwise} \end{cases}$

- $X^{cl}_{O_i} = \begin{cases} 1, & \text{if } i\text{-element of the set } O \text{ belongs to the level of the core} T^{cl} \\ 0, & \text{otherwise} \end{cases}$

To implement the distributed structure of the information system of a multifunctional measuring complex, the following conditions and restrictions are introduced: each element belongs to only one level

$$\sum_{i=1}^{n_{sw}} (X^{el}_{O_i} + X^{fI}_{O_i} + X^{cl}_{O_i}) = 1$$

A workstation can be connected to only one switch $SW_j$

$$\sum_{j=1}^{n_{sw}} X^{el}_{WS_iSW_j} X^{el}_{SW_j} = 1, i = 1, n_{ws}$$

where

$$X^{el}_{WS_iSW_j} = \begin{cases} 1, & \text{if the workstation } WS_i \text{ is connected to the switch } SW_j. \\ 0, & \text{otherwise} \end{cases}$$

Connection of active network equipment of the level of boundary computing to the level of fog computing taking into account the presence of backup connections

$$\sum_{j=1}^{n_{sw}} X^{el}_{SW_iSW_j} X^{el}_{SW_i} X^{fI}_{SW_j} \geq 1, i = 1, n_{sw}$$

where

$$X^{el}_{SW_iSW_j} = \begin{cases} 1, & \text{if the switch } SW_i \text{ is connected to the switch } SW_j. \\ 0, & \text{otherwise} \end{cases}$$

Servers can be connected to active network equipment at fog computing level.

$$\sum_{j=1}^{n_{sw}} X^{fI}_{SR_iSW_j} X^{fI}_{SW_j} = 1, i = 1, n_{sw}$$

where

$$X^{fI}_{SR_iSW_j} = \begin{cases} 1, & \text{if the server } SR_i \text{ is connected to the switch } SW_j. \\ 0, & \text{otherwise} \end{cases}$$

Connection of active network equipment of the level of fog computing to the level of cloud computing


\[ \sum_{j=1}^{n_{sw}} x_{sw_{i}sw_{j}} x_{sw_{i}}^{f} x_{sw_{j}}^{c} \geq 1, i = 1, n_{sw} \]

Servers can be connected to active cloud computing network equipment.

\[ \sum_{j=1}^{n_{sw}} x_{sr_{i}sw_{j}} x_{sw_{j}}^{c} = 1, i = 1, k_{s} \]

Within the physical topology, the connectivity of the computer network must be ensured.

\[ \text{conn}(o_{i}, o_{j}) = 1, o_{i} \in WS \cup SR \cup SS, i \neq j \]

where \( \text{conn}(o_{i}, o_{j}) \) – a function that is equal to one if there is a path between the network objects \( o_{i} \) and \( o_{j} \).

The effectiveness of the implemented topological structure is determined by the following formula

\[ FT(x_{sw_{i}sw_{j}}^{f}, x_{sw_{i}}^{c}, x_{ws_{i}sw_{j}}^{f}, x_{sr_{i}sw_{j}}^{c}, x_{ss_{i}sw_{j}}^{c}, x_{sw_{i}sw_{j}}^{el}, x_{sw_{i}}^{f}, x_{sw_{j}}^{c}) \rightarrow \min \max \]

3. Algorithm solutions

Given the relatively large dimension of certain information-measuring systems, the process of designing them requires solving an NP-complete problem and therefore it is advisable to use heuristic algorithms, in particular, from the family of genetic algorithms. The advantages of genetic algorithms are that they work with codes that represent a formalized view of a set of parameters that are arguments of the objective function; during the implementation of the search procedure, the genetic algorithm simultaneously processes several points of the search space, which makes it possible to overcome the danger of falling into the local extremum of the multimodal objective function; in the process of work do not use any additional information; to generate new points of the search space at the same time, the genetic algorithm uses both probabilistic and deterministic rules, which gives a much larger effect [3–8].

The basic idea of a genetic algorithm is to create a population of individuals, each of which is represented as a chromosome. Any chromosome is a possible solution to the optimization problem under consideration. To find the best solutions, only the value of the objective function (or fitness function) is used. The value of the individual's fitness function shows how well the individual described by this chromosome is suitable for solving the problem. The chromosome consists of a finite number of genes, representing the genotype of the object, i.e. the totality of its hereditary traits. The process of evolutionary search is conducted only at the level of the genotype. The main biological operators are applied to the population: crosses, mutations, inversions, etc. The population is constantly updated by generating new individuals and destroying old ones, and each new population becomes better and depends only on the previous one.

The use of a genetic algorithm requires the development of a solution coding method. For implementation, a chromosome structure was proposed which represents a string consisting of genes that determine the interconnection of network infrastructure objects at the cloud, fog and boundary computing levels.

Each chromosome is estimated by the measure of its “fitness” (fitness-function). The most adapted individuals have a greater opportunity to participate in the reproduction of offspring. Proportional selection assigns to each \( i \)-th chromosome a probability \( P(i) \) equal to the ratio of its adaptability to the total fitness of the population. In this case, a multipoint crossover uses a break point, which is the boundary between adjacent array elements (i.e., a node is randomly selected). When choosing from which parent the descendant will take the next gene, preference is given to the most adapted representative of the population.

The genetic algorithm includes the following steps:

- an initial population of \( n \) chromosomes is determined, each of which represents a solution, that is, the objects of the network infrastructure are randomly distributed into levels;
- the coefficient of fitness for each chromosome is calculated and, in accordance with it, a new population of the most adapted individuals is created;
• for each population, we recalculate the distribution of network objects using the best chromosome from the current population;
• we receive new network infrastructure and the algorithm continues to work until it is, the stopping criterion is reached.

As a result of the use of genetic operators, a chromosome is obtained, which is a possible solution. During the study of the genetic algorithm for solving the problem of choosing the optimal variant in distributed information-measuring systems, a number of approaches and modifications of the classical algorithm were used, which implied, in particular, a variable mutation in which the probability of mutation varies depending on the needs of the algorithm (Figure 1).

![Figure 1. Experimental results comparing the efficiency of constant and variable mutation.](image)

4. Conclusion
The main conclusion is the possibility of using a multi-level representation of the topological structure to determine the optimal architecture of the distributed information system of a multifunctional measuring complex based on cloud, fog and boundary calculations. Considering the scale of modern distributed multifunctional measuring systems when building intelligent power supply networks or monitoring systems for transporting gas or oil products, it becomes expedient to use a genetic algorithm to determine a rational version of the computing network infrastructure. Also, genetic algorithms should be used as a tool for managing the SDN network of the measuring complex.

As a result, the distributed information system of a multifunctional measuring complex can be described by a mathematical model within the framework of a multi-level topological computational structure network combining cloud, fog and boundary computing complexes into a single network infrastructure. This allows building a modern unified network infrastructure of a multifunctional measuring complex as part of the industrial IoT concept, which in turn ensures the virtualization of production and technological functions both within monitoring and within management, the ability to analyze large amounts of data in order to build a model to improve customer experience and reducing its costs, building new efficient business models, as well as accelerating the response to dynamically changing market requirements and overall trend to reduce product life cycle. On the other hand, the unification and standardization of technological approaches for building a multifunctional measuring complex based on cloud, fog and boundary computing models facilitate management and make the infrastructure accessible to small and medium businesses within the general direction of digitalization.
Acknowledgements
The study results are obtained with the support of the project #2.7193.2017/8.9 “Development of scientific bases of design, identification and diagnosis systems for highly accurate positioning with application of the methodology of inverse problems of electrical engineering”, carried out within the framework of the base part of State job.

References
[1] Vorobyev S P 2009 *Issues of modern science and practice* **8** 131–143
[2] Vorobyev S P 2016 *Journal of Engineering and Applied Sciences* **11** (6) 1243–1247
[3] Back T, Fogel D B and Michalewicz Z 1997 *Handbook of Evolutionary Computation* (New York: Oxford University Press)
[4] Davis L 1987 *Genetic Algorithms and Simulated Annealing* (London: Pitman)
[5] Kureychik V M 1988 *Genetic Algorithms* (Taganrog: TRTU)
[6] Kureychik V M 1998 *Genetic search methods* (Taganrog: TRTU)
[7] Kureychik V M 2002 *Genetic algorithms and their application* (Taganrog: TRTU)
[8] Goldberg D E 1989 *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison: Wesley Publishing Company)