Models and Methods for Determining Application Performance Estimates in Distributed Structures

Viktor Grechaninov[1], Oleksandr Khoshaba[1], Hennadii Hulak[2], Yuliia Zhdanova[2], and Iryna Melnyk[2]

[1] *Institute of Mathematical Machines and Systems, 42 Glushkova ave., Kyiv, 03187, Ukraine*
[2] *Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska str., Kyiv, 04053, Ukraine*

Abstract

The method of evaluating the operation of service programs at nodes of the distribution system is proposed, which consists in the use of a predictive control model, is proposed. A general analysis was carried out between the use of one of the existing and reviewed methods of determining the evaluations of nodes in distributed systems based on models with consideration and prognostic control, where their main characteristics are considered. Features are presented of using PID regulators based on models of dynamic systems are analyzed. The experimental results of the proposed method of determining performance evaluations of applied applications in distributed structures based on the predictive control model are described.

Keywords

Predictive control model, distributed systems, proportional-integral-differential, PID, regulator, automated control system.

1. Introduction

Currently, the use of applications in distributed structures rather than monolithic ones is gaining more and more popularity. The need to use distributed applications based on modern software and technical complexes is caused not only by the spread of large data centers of cloud providers, but also by the increasingly frequent emergence of corporate networks in commercial and budget organizations. In this connection, there are problems related to the assessment of efficiency of the use of applications and security in the work of computing and network resources of the corporate network [1]. The most common problems include changing the estomates of the efficiency and productivity of nodes of a distributed system as a result of its modernization with software and technical means, installing patches on the operating systems of nodes, adding new ones, as well as studying the software code of existing service applications, implementing scripts for performing work on deploying information structures, putting additional nodes into operation, etc. [2, 3].

In the sphere of security of systems, there is a problem of identifying unauthorized actions by intruders. Such actions lead to a change in the evaluations of the efficiency and productivity of the nodes in relation to the reference (previously measured) values.

Using model-based assessments in existing industrial or IT infrastructures to configure auto-scaling of corporate network nodes also results in significant cost savings.

2. Research Purpose

The purpose of the paper is to study the functioning of service applications on the nodes of a distributed structure using dynamic systems and the Model of Predictive Control (MPC). The study includes obtaining estimates of service application performance in a distributed system based on load effects by means of a series of requests to the object of study, which are formed using reference trajectories.

CPITS-2022: Cybersecurity Providing in Information and Telecommunication Systems, October 13, 2022, Kyiv, Ukraine
EMAIL: grechaninov@nas.gov.ua (V. Grechaninov); oleksandr.khoshaba@gmail.com (O. Khoshaba); h.hulak@kubg.edu.ua (H. Hulak); y.zhdanova@kubg.edu.ua (Y. Zhdanova); iy.melnyk@kubg.edu.ua (I. Melnyk)
ORCID: 0000-0001-6268-3204 (V. Grechaninov); 0000-0001-5375-6280 (O. Khoshaba); 0000-0001-9131-9233 (H. Hulak); 0000-0002-9277-4972 (Y. Zhdanova); 0000-0001-6041-6145 (I. Melnyk)

3. Method of Determining Estimates on Nodes in Parallel and Distributed Structures

Method of solving the problems of determining the estimates of the use of this function on nodes in parallel and distributed structures

In order to solve the problems of determining the estimates of node performance, models of parallel and distributed structures based on reference algorithms for problem solving are proposed [1–4]. In this regard, the determination of node performance estimates, which is based on the model of a parallel system, uses a reference sequential algorithm for solving some problem *A* by an application in time *T* [2, 4]. In this case, the acceleration estimate is used, which is defined as:

$$S = \frac{T_0}{T}. \quad (1)$$

where T_0 is solution time of reference problem *A* by the application on one device (node) using the fastest sequential algorithm.

Acceleration *S* shows how many times the application's problem solving time can be reduced by using a parallel structure [3].

The next assessment of the performance of applications on nodes in parallel structures is efficiency, which is defined [1, 3] as:

$$E = \frac{S}{n} = \frac{T_0}{nT}. \quad (2)$$

The described model of using this function on a parallel structure is simple to calculate estimates. However, in this model, the value of their estimates is determined only after the completion of the task, when the total time *T* is known. In addition, the model requires knowledge of the time of solving the problem by the best from a set of sequential algorithms T_0 on one of the nodes on the parallel structure.

Another approach to using similar models for job evaluations employs distributed structures in connection with two aspects. Firstly, nodes in distributed structures due to their heterogeneity may have different values of application performance estimates as these nodes have different computing resources. Secondly, nodes in distributed structures may be unavailable at certain time intervals during the solution of the entire problem.

Let us consider one of the models for applications to distributed structures,which is characterized by the presence of a schedule and is defined as:

$$h_i(t): R^+ \rightarrow \{0,1\}. \quad (3)$$

At the same time, $h_i(t) = 1$ if the application located on the node at time t is available to solve task *A*. Otherwise, if $h_i(t) = 0$, then the application on the node at time t is unavailable.

Performance estimates for applications located on a distributed structure take other ratios. Thus, for a model with a schedule of a distributed structure, the efficiency assessment takes the form [5]:

$$E_t = \frac{\underline{T}(A)}{T(A)}. \quad (4)$$

where *T*(*A*) is the reference time for solving the task *A* by the application.

The reference time $\underline{T}(A)$ is the time of solving task A by the ith application on the node using the fastest sequential algorithm, where $\underline{T}(A) > 0$.

Also, for applications located on a distributed structure, such additional estimates as [5, 6] are introduced: the performance of the reference system and the complexity of the task. The calculation of the reference performance of the system $\pi(A,t)$ [5–7] is necessary for the calculation of *T*(*A*).

The reference performance of the structure $\pi_i(A,t)$ is called [7] the performance of the i^{th} node of the distributed system when solving the problem *A*:

$$\pi_i(A) = \frac{L(A)}{\underline{T_i}(A)}. \quad (5)$$

where *L*(*A*) is a function of the complexity of the task execution by the application on the distributed structure.

The task complexity function *L*(*A*) is defined on some set of tasks Λ and expresses a priori knowledge of its complexities:

$$L(A): \Lambda \rightarrow R^+. \quad (6)$$

For example, the complexity function of the performance of the task *L*(*A*) by an application on a distributed structure can include an estimate of the number of elementary operations, which are calculated using the complexity theory [8–10] . To solve the problems of determining the evaluations of node performance in parallel and distributed structures, a model with a schedule is introduced, which describes the concept of the reference performance of the structure as the sum of the reference performances of the nodes [7]:

$$\pi(A,t) = \sum_{i=0}^{n} \pi_i(A)h_i(t) = \left(\vec{\pi}(A), \vec{h}(t)\right), \quad (7)$$

where with full availability of applications on nodes $h_i(t) \equiv 1$ in the process of solving the problem and with the same reference performance, where $\pi_i(A) = \pi_0$, the reference performance of the distributed structure will coincide with the reference performance of the parallel structure as the condition $n \cdot \pi_0$ will be fulfilled.

At the same time, the model with the schedule $\mathfrak{R}$ for solving problems from the set Λ is called the set:

$$R = < \vec{\pi}(A), \vec{h}(t) >. \quad (8)$$

For a model with the schedule $\mathfrak{R}$, the reference time for solving problem A is called the value $T(A)$ [5–7], which is determined by the following relation:

$$\underline{T}(A) = t \colon \int_{t=0}^{t} \pi(A,\tau)d\tau. \quad (9)$$

In the general case, the acceleration indicators for the parallel structure are defined as the ratio of the time of solving the problem by the application at one node to the time of solving the problem on the whole system based on the ratio (1). In a distributed structure with a schedule (8), the resources of the nodes may be different. Therefore, it is incorrect to enter the concept of acceleration to the applications in the distributed structure, since it is not clear about which node is to calculate the acceleration parameter itself. On this basis, a more general concept of relative acceleration is introduced, which is determined as follows:

$$S(R_1, R_2) = \frac{T_1}{T_2}. \quad (10)$$

The ratio (10) shows the evaluation of the acceleration S for the model with the schedule R_1 relative to the other system R_2 as the ratio of their time of solving the problem A on the applications in the distributed structure. Also, for a schedule model, it is customary to perform the acceleration evaluation S for each node of the distributed structure as follows:

$$S_i = \frac{T_i}{T}. \quad (11)$$

The ratio (11) shows the estimate of acceleration Si for the i^{th} application on the node as the ratio of the reference time of solving problem A to the time of solving this problem in a distributed structure based on a schedule model.

4. Method of Solving the Problem Determining the Performance

The proposed method of solving the problem of determining the performance of applications in distributed structures is based on the following aspects. Firstly, such a model of research as a distributed structure is a dynamic system. Therefore, evaluations of the study of an object (service applications) and the regulator of load effects in distributed structures (Fig. 1) can be obtained in this way:

- To determine the main characteristics of the object of study (k, θ, τ) it is necessary to perform its identification according to the models (13–17).
- To determine the main characteristics of the regulator, the calculations according to the model (18) are performed.

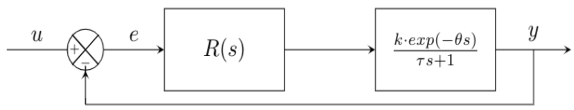

Figure 1: General scheme for determining the estimate of the object of research and regulator of load effects in distributed structures

Secondly, the proposed method uses the reference function of the load effect on the object of research, which can be represented by both linear and non-linear model.

The comparative characteristics of the existing and proposed methods of determining the evaluations of application functions at the nodes in the distributed structure are presented in Table 1, which shows the main distinctive features.

The method of control with predicting models is one of the modern methods of management theory, which is most effectively used in management of technological processes. Although this method has become widely used since the early 80's of the 20th century, it has today been improved and practiced in classical management with negative feedback. This method is based on predicting the behavior of the control object on different types of input influences. Feedback in such control systems is used to adjust inaccuracies related to external obstacles and inaccuracy of the mathematical model in relation to the control object. To do this, a regulator is created that uses the empirical model of the control object to predict its further behavior.

Table 1
Comparative characteristics of the existing and proposed method of determining the evaluations of applications at nodes in a distributed structure

Characteristics	Existing method	Proposed method
The nature of the submission of a distributed system	In the form of a static system	In the form of a dynamic system
Models of research	Schedule model	Model of prognostic control
Input parameters of a model	Time to solve a problem on a desktop application	Time to process the request to the service program
Object of research	Desktop programs are located at the distributed system nodes	Service programs are located at the distributed system nodes
Output parameters of a model	Efficiency, acceleration, productivity, complexity	Parameters of the model of the object of study, parameters of control regulator
Load effect tools	A set of consecutive algorithms based on reference algorithm	A set of consecutive and parallel series of requests based on the reference trajectory

In many cases, the problems listed above and related to the evaluation of the performance of applications on nodes in distributed structures, can be effectively solved with the help of some set of dynamic optimization algorithms [11]. In the field of management of service applications on nodes of distributed structures, using dynamic optimization methods it is possible to maintain a given level according to the reference trajectory of load effects. Such a level may have a non-linear function of the trajectory according to which load effects on the research object are performed. For this purpose, with the help of the optimality function used, the best management of the process of load effect on service applications of distributed structures is carried out, which is reduced to the maximization or minimization in time of the selected criterion.

Dynamic optimization, which is used in the proposed method of solving the problem of determining application performance estimates in distributed structures, effectively uses systems of differential algebraic equations (DEA) to perform numerical solutions. One of the most popular practical implementations of dynamic optimization includes the model predictive control (MPC) method (Fig. 2). The main purpose of using MPC is to minimize the difference between the given value of the controlled variable and the predictions of the model.

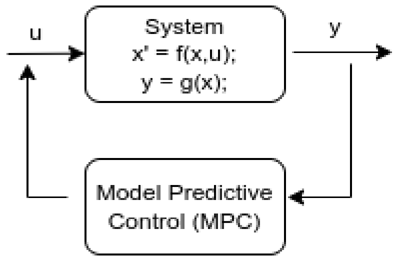

Figure 2: Practical implementation of dynamic optimization of the research object (service applications in a cluster structure)

The model of the control object is usually chosen to be linear, but in this paper we will show the operation of control with predictive models on a non-linear model (Fig. 5). In the case of determining the estimates of applications in distributed systems, a given non-linear model of the control object (service applications) is set as the trajectory of the reference effect of the load (u). As a result of this formalization of the subject area, we will get the following model of predictive control in the form of the variable y (Fig. 2):

$$\tau\frac{dy}{dt} = -y + ku, 0 \le u \le 500. \quad (12)$$

Management models with the following transferable functions are defined for the research object (service applications in distributed structures):

$$\frac{y(s)}{u(s)} = \frac{k\, exp(-\theta s)}{\tau s + 1}. \quad (13)$$

$$\frac{y(s)}{u(s)} = \frac{k\, \omega^2\, exp(-\theta s)}{s^2 + 2\xi\omega s + \omega^2}. \quad (14)$$

$$\frac{y(s)}{u(s)} = \frac{k\, exp(-\theta s)}{s}. \quad (15)$$

$$\frac{y(s)}{u(s)} = \frac{k\,exp(-\theta s)}{s^2}. \qquad (16)$$

$$\frac{y(s)}{u(s)} = \frac{k\,exp(-\theta s)}{s(\tau s + 1)}. \qquad (17)$$

Based on the models described above, such parameters as transmission coefficient (k), delay time (θ), and time constant (τ) are most often determined.

Thus, as a result of the load effects on the research object, the reaction of the system (distributed structure) in the form of measurements (request processing) is marked (y_t). The model built on the basis of the scheme of Fig. 1 with the calculated parameters of the regulator (18, Fig. 3) and the research object (13, Fig. 1) is denoted as y.

5. Peculiarities of using PID Regulators based on Models of Dynamic Systems

Models of dynamic systems often have a delay link in their structure, the cause of which is the peculiarities of technological processes. Also, in dynamic systems, the impact on the research object and the reaction of the research object is a function of time (12). Therefore, the reaction to the impact of the research object can be determined both by the current and previous values of the impact on it. As a result, the dynamic system has inertia.

At the same time, during the development of such systems with a delay, rather efficient PIDs and some other regulators with a special structure are mostly used. Also the advantage in use is given to the PID family of regulators due to their simplicity, efficiency and prevalence. It is known [12] that earlier in more than 90% of cases when using technological process control systems in dynamic systems, PID regulators were used (Fig. 3).

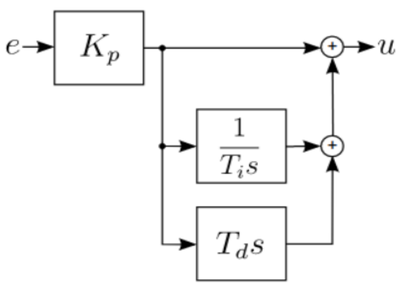

Figure 3: Scheme of PID regulator

During work with PID regulators, more than a hundred options for determining their settings were created [13–16]. This number is determined by the difference in the reference models of the dynamics of technological processes, the quality criteria of transient processes, the conditions of their application, the accuracy and reliability of the algorithms for calculating and optimizing their parameters. The basis of the use of regulators in dynamic structures is the study of object control systems, in which the delay time for a change in the controlling effect takes most of the time for the reaction of the object of study.

In the classical theory of automatic control, the regulator structure is selected from the control object model, for example (13)–(17). At the same time, difficult control objects require the use of complex regulators. However, in practice, in the vast majority of cases, regulation is reduced to the use of PID regulators according to the model:

$$u = K_p(1 + \frac{1}{T_i s} + T_d s). \qquad (18)$$

PID controllers do not always provide the required quality of regulation, but due to the simplicity of their structure and a large number of theoretical and practical methods of their adjustment, PID controllers are the main ones in practical application.

Several types of criteria are used to assess the quality of transient processes: direct, integral and frequency. The most important direct indicators are the maximum deviation module |ymax| and adjustment time Tp. Among others, integral and frequency indicators are most often used. For example, among frequency indicators of quality (MS), the sensitivity function [12] is very often used, which is defined as follows:

$$M_S = \max_{\omega} \left| \frac{1}{P(j \cdot \omega) \cdot C(j \cdot \omega) + 1} \right| \qquad (19)$$

where $P(s)$ and $C(s)$ are transmission functions of the control object and regulator; ω is circular frequency; j is an imaginary unit.

6. Experimental Research

Experimental studies of the models and the proposed method of determining application performance estimates in distributed structures were conducted on the basis of the system shown in Fig. 4.

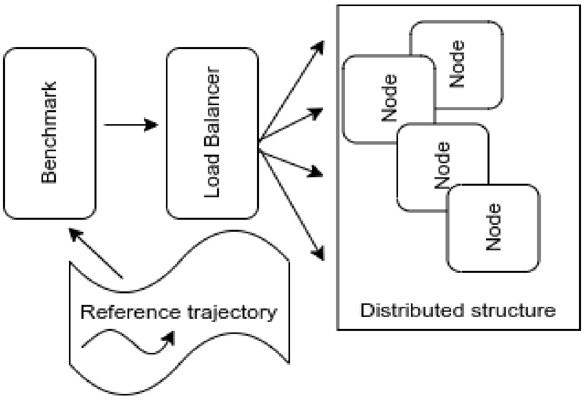

Figure 4: The system of conducting experimental research

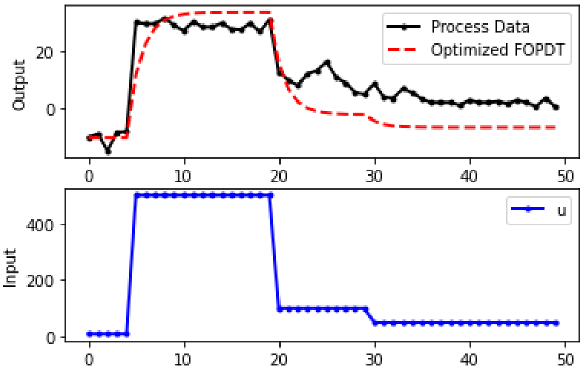

Figure 5: Results of the analysis of the use of the predictive control model

Fig. 4 presents a general scheme of the system for conducting experimental research and data collection, which consists of the following main components: a benchmark, a load balancer and nodes of a distributed (cluster) structure.

Let's consider the main tasks performed by each of the main components. The benchmark, according to the reference trajectory of the load impact *u* (Fig. 2), ensured the generation of a series of requests during certain time intervals for the load balancer. The load balancer, in turn, performed traffic redirection (a series of requests) according to the algorithm of uniform distribution to the nodes of the cluster structure. After that, the benchmark received a series of requests from the load balancer and recorded the response time (RT). As a result, an assessment of the load effects (LE) on the service programs of nodes of the cluster structure was used, which was performed as follows:

$$L_E = k \cdot ln\, ln \frac{RT_i}{PT_i} \qquad (20)$$

where RT_i is time of processing and transportation of a series of requests to the research object; PT_i is the period between the generation of successive series of requests for the research object; k is a correction factor equal to 10.

Assessment of load effects (LE) corresponds to the notation y, which is shown in Fig. 2.

The experimental data obtained in this way were analyzed. The results of the analysis of the use of the predictive control model are shown in Fig. 5.

When performing experimental studies, the following task was set: based on the given values of the reference effects (*u*), to obtain the values of the estimates of the load effects. Next, it was necessary to build and optimize the model based on the transmission function (13) and the PID regulator.

The assessment of load effects (20) is shown in the graph of Fig. 5 and corresponds to the inscription Process Data. In the course of experimental studies, the estimate (20) was obtained as a reaction of the system (Fig. 2) to the load effects u (the lower part of the graph in Fig. 5). The values obtained as a result of optimization are shown in Fig. 5 and labeled as Optimized FOPDT (upper part of the graph).

According to the analysis shown in Fig. 5 for the research object (service applications) the transfer function was determined according to (13) with the parameters shown in Table 2. The root mean square error based on the found parameters (Table 2) was 7589.8.

Table 2
Parameters of the research object mode

Parameters of the model	Values of the parameters (initial values)
Transmission coefficient (*k*)	0.06
Delay time (ϑ)	2.32
Time constant (τ)	0

The parameters of the control regulator were also determined (Table 3) using the efficiency indicator of the optimization process:

$$E = \frac{R_0}{R} \cdot 100\,\% . \qquad (21)$$

where R_0 is the estimate of the root mean square error obtained as a result of optimization of the parameters of the transition function of the research object and the PID of the regulator; R is initially obtained estimate of the root mean square error as a result of determining the parameters of the transition function of the research object and the PID regulator.

In this work, the use of the method of control management (MPC, Fig. 2) made it possible to increase the accuracy of the predictive model by 70.22%.

Table 3
Parameters of the control regulator

Parameters of the regulator	Values of parameters of the regulator
Proportional gain	18.3
Integral gain	2.3
Derivative gain	0
Initial root mean square error	7589.8
The final root mean square error	2259.9

The following software was used during the experimental research: the Ubuntu version 21.04 operating system, a cluster structure based on MicroK8s and Docker containers.

7. Conclusions

So the presentation of distributed structures as dynamic systems allows:

- To receive information about existing changes (for improvement or deterioration) of qualitative and quantitative assessments of the main characteristics of applications on nodes contained in distributed (cluster) structures in connection with their purposeful (sanctioned) changes in the software code or changes in the software settings environment (operating systems, system utilities).
- Create an automatic control system that works in real time by determining the optimal characteristics of the transfer function of the model of the research object and the PID regulator.
- On the basis of the obtained automatic control system, determine the margin of safety of the distributed (cluster) structure due to the appearance of unaccounted for external or internal excitations (noises) on the research objects or the software environment.

8. References

[1] M. Vladymyrenko, et al., Analysis of Implementation Results of the Distributed Access Control System, in VI International Scientific and Practical Conference Problems of Infocommunications. Science and Technology, 2019, pp. 39–44. doi: 10.1109/PICST47496.2019.9061376.

[2] V. Buriachok, V. Sokolov, P. Skladannyi, Security Rating Metrics for Distributed Wireless Systems, in 8th International Conference on "Mathematics. Information Technologies. Education," vol. 2386, 2019, pp. 222–233.

[3] V. Astapenya, et al., Analysis of Ways and Methods of Increasing the Availability of Information in Distributed Information Systems, in 8th International Conference on Problems of Infocommunications, Science and Technology, 2021, pp. 174–178. doi: 10.1109/PICST54195.2021.9772161.

[4] A. Afanasiev, M. Posypkin, A. Khritankov, Analytical Model for Evaluating the Performance of Distributed Systems, Software Products and Systems, vol. 4, 2009, pp. 60–64.

[5] A. Khritankov, Mathematical Model of Performance Characteristics of Distributed Computing Systems, in 50th scientific conference of MIPT, 2007, pp. 86–88.

[6] A. Khritankov, On Performance Characteristics of Distributed Systems, in 51st scientific conference of MIPT, 2008.

[7] A. Khritankov, Models and algorithms for load distribution. Algorithms based on QS networks, Information technologies and computing systems, vol. 3, 2009, pp. 33–48.

[8] M. Posypkin, A. Khritankov, On the Concept of Acceleration and Efficiency in Distributed Systems, in All-Russian Scientific Conference "Scientific Service on the Internet: Solving Big Problems," 2008, pp. 149–155.

[9] A. Khritankov, Evaluation of the Efficiency of Distributed Systems in Solving Problems of Variable Size, Scientific and technical bulletin of St. Petersburg State University ITMO, vol. 2, 2010, pp. 66–71.

[10] M. Posypkin, A. Khritankov, On the Concept of Performance in Distributed Computing

Systems, in ISA RAS, vol. 32, 2008, pp. 26–32.
[11] V. Grechaninov, et al., Decentralized Access Demarcation System Construction in Situational Center Network, in Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3188, no. 2, 2022, pp. 197–206.
[12] J. L. Balcazar, J. Dıaz, J. Gabarro, Structural Complexity II, vol. 22. EATCS Monographs on Theoretical Computer Science, 1990. doi: 10.1007/978-3-642-75357-2.
[13] J. L. Balcazar, J. Dıaz, J. Gabarro, Structural Complexity I, 2nd Ed. Theoretical Computer Science, An EATCS Series, 1995. doi: 10.1007/978-3-642-79235-9.
[14] O. Goldreich. P, NP, and NP-Completeness: The Basics of Complexity Theory. Cambridge University Press, 2010.
[15] K. J. Åström, T. Hägglund, Advanced PID control, USA, Instrumentation, Systems, and Automation Society, 2006.
[16] A. O'Dwyer, Handbook of PI and PID Controller Tuning Rules, London, Imperial Colledge Press, 2006.