OPTIMIZATION OF CLOSED STOCHASTIC NETWORKS

V. M. Vishnevskii and Z. L. Kruglyi

UDC 62-505:65.012.122

An algorithm of optimization of closed queuing networks is proposed that have the property of local balance. The capacity, the load, and the average length of the queue are studied as a function of the service rate. This algorithm is used for optimizing the capacity of computer systems and networks.

1. Introduction

Queuing networks (QUNT) are widely used in the analysis of computer systems and networks [1-7]. However, the methods of optimization of QUNT needed for the designing of systems have not yet been sufficiently developed.

At present there exist approximate methods of optimization of closed networks [8, 9], as well as exact methods of optimization of network models of multiprogram computers [10, 12]. In this paper we consider an exact algorithm of optimization of homogeneous closed QUNT of any structure, i.e., we study the capacity, the load, and the average queue length as a function of the service rate. For network optimization we obtain a system of equations that does not contain partial derivatives. By such an approach it is possible to extend the domain of application of optimization methods to networks, and to increase the computational effectiveness of the algorithms.

2. The Capacity as a Function of the Service Rate

We shall consider separable or locally balanced networks (LBN) [11]. Suppose that a closed LBN contains M nodes and N calls which circulate between the former. The route of the calls along the network is specified by a stochastic indecomposable matrix $P = \|P_{ij}\|$, where P_{ij} is the probability that after ending the service at the i-th node, a call will arrive at the the j-th node $\left(i, i=\overline{1, M}; 0 \leqslant P_{ij} \leqslant 1; \sum_{i=1}^{M} P_{ij} = 1\right)$. We shall assume that the rate of service at the i-th node ($i = \overline{1, M}$) depends on the number of calls in the queue and that it can be written in the form $\mu_i(k) = \beta_i(k)\mu_i$. For example, for the case often encountered in practice, when the i-th node is a multiserver queuing system with n_i servers and a distribution function of the service time of calls in each server equal to $F_i(t) = 1 - \exp\{-\mu_i t\}$, we have $\mu_i(k) = \beta_i(k)\mu_i$, where

$$\beta_i(k) = \begin{cases} k, & \text{for } k \leqslant n_i, \\ n_i, & \text{for } k > n_i. \end{cases}$$

The capacity and the average number of occupied servers of the i-th node* are expressed [11] by the formulas

$$\lambda_i(N) = e_i G(N-1)/G(N), \tag{1}$$

$$U_i(N) = x_i G(N-1)/G(N), \tag{2}$$

$$x_i = e_i/\mu_i, \quad \bar{e} = \bar{e}P, \tag{3}$$

where G(n) is the normalizing constant of a closed network in which n calls are circulating ($n = \overline{1, N}$); $\bar{e} = \{e_1, e_2, \ldots, e_M\}$ is the vector of the relative rates of the flows specified by the system of equations (3); x_i is the relative occupation factor.

*This refers to the case that the i-th node is a multiserver queuing system.

Moscow. Minsk. Translated from Avtomatika i Telemekhanika, No. 2, pp. 41-53, February, 1987. Original article submitted December 19, 1985.

0005-1179/87/4802-0173$12.50

The average length of the queue at the i-th node (taking into account the calls that are being served) can be expressed as follows:

$$L_i(N) = \sum_{n=1}^{N} nP_i(n), \tag{4}$$

where $P_i(n)$ is the marginal probability that the number of calls at the i-th node is equal to n, $n = \overline{0, N}$.

At first let us study $\lambda_i(N)$, $U_i(N)$, and $L_i(N)$ as a function of the μ_i (or x_i) of its "own" node.

THEOREM 1. If the rate of service at each node of a closed LBN with N calls and M nodes is a nondecreasing function of the number of calls in the queue, i.e., $\mu_i(k+1) \geq \mu_i(k)$, then the average number of occupied servers and the average queue length will be monotonically decreasing (increasing) functions, whereas the capacity is a monotonically increasing (decreasing) function, of the service rate (the relative occupation factor) of its "own" node:

$$\partial U_i(N)/\partial \mu_i = -\frac{1}{\mu_i} U_i(N)[1 + L_i(N-1) - L_i(N)], \tag{5}$$

$$\partial L_i(N)/\partial \mu_i = -\frac{1}{\mu_i} D_i(N), \tag{6}$$

$$\partial \lambda_i(N)/\partial \mu_i = \frac{1}{\mu_i} \lambda_i(N)[L_i(N) - L_i(N-1)], \tag{7}$$

$$\partial U_i(N)/\partial x_i = \frac{1}{x_i} U_i(N)[1 + L_i(N-1) - L_i(N)], \tag{8}$$

$$\partial L_i(N)/\partial x_i = \frac{1}{x_i} D_i(N), \tag{9}$$

$$\partial \lambda_i(N)/\partial x_i = -\frac{1}{x_i} \lambda_i(N)[L_i(N) - L_i(N-1)], \quad i = \overline{1, M}, \tag{10}$$

where $D_i(N)$ is the variance of the number of calls at the i-th node, and $L_i(k)$ is the average queue length at the i-th node for a network with k calls, $k = \overline{1, N}$.

Theorem 1 is proved in Appendix 1.

For LBN it is possible to prove the following theorem which expresses the relationship between the characteristics of a node when the service parameters of another node of the network vary.

THEOREM 2. If the rate of service at each node of a closed LBN with N calls and M nodes is a nondecreasing function of the number of calls in the queue, then the capacity, the average number of occupied servers, and the average queue length will be monotonically increasing (decreasing) functions of the service rate (the relative occupation factor) of "another" node:

$$\partial U_i(N)/\partial \mu_j = \frac{1}{\mu_j} U_i(N)[L_j(N) - L_j(N-1)], \tag{11}$$

$$0 \leqslant \frac{\partial L_i(N)}{\partial \mu_j} \leqslant \frac{1}{\mu_j} D_j(N), \tag{12}$$

$$\frac{\partial \lambda_i(N)}{\partial \mu_j} = \frac{1}{\mu_j} \lambda_i(N)[L_j(N) - L_j(N-1)], \tag{13}$$

$$\frac{\partial U_i(N)}{\partial x_j} = -\frac{1}{x_j} U_i(N)[L_j(N) - L_j(N-1)], \tag{14}$$

$$-\frac{1}{x_j} D_j(N) \leqslant \frac{\partial L_i(N)}{\partial x_j} \leqslant 0, \tag{15}$$

$$\frac{\partial \lambda_i(N)}{\partial x_j} = -\frac{1}{x_j} \lambda_i(N)[L_j(N) - L_j(N-1)], \quad i, j = \overline{1, M}, i \neq j. \tag{16}$$

Theorem 2 is proved in Appendix 2.

To study the capacity as a function of the service rate is important also for constructing iteration algorithms of design of LBN [6, 7].

3. Statement of Problem of Optimization of Closed Networks

The cost of a network will be defined as follows:

$$F=\sum_{i=1}^{M} c_i\mu_i^{a_i},$$

where c_i is a cost coefficient, μ_i the rate of service at the i-th node, and a_i a nonlinearity factor.

Since the capacities of the nodes in an exponential network are proportional to one another, it follows that the capacity of a network can be defined in terms of the capacity of one of the nodes of the network, i.e., $\lambda = \lambda_i/\alpha_i$, where the α_i are the relative average rates of utilization of the network nodes.

The problem of optimizing a closed network can have one of the following formulations.

Formulation 1. Find

$$\max \lambda = \frac{1}{\alpha_1} e_1 G(N-1)/G(N) \tag{17}$$

under the constraints

$$F=\sum_{i=1}^{M} c_i\mu_i^{a_i} \leqslant S, \quad \bar{\mu}>0. \tag{18}$$

Formulation 2. Find

$$\min F=\sum_{i=1}^{M} c_i\mu_i^{a_i} \tag{19}$$

under the constraints

$$\lambda=\frac{1}{\alpha_1} e_1 G(N-1)/G(N) \geqslant \Lambda, \quad \bar{\mu}>0.$$

The solution is sought on the set of values of the service rates. In this formulation it is assumed that the speeds of the servers are continuous variables, whereas in actual fact these variables are discrete. Since in the case of a large dimension the solving of a discrete optimization problem is laborious from a computational point of view, we shall at first solve the optimization problem in terms of nonlinear programming. After that we can have discrete search (of smaller dimension) together with search for a continuous optimum.

In considering a fairly large class of cost functions, it was proved in [12] that any local maximum of the problem in Formulation 1 will be also a global maximum, whereas any local minimum of the problem in Formulation 2 is also a global minimum.

4. Maximization of Network Capacity with a Cost not Exceeding the Assigned Value

An optimal solution of (17) will be sought by the method of undetermined Lagrange multipliers.

Let us construct the Lagrange function $\dot{Q} = \lambda + \gamma(F - S)$, where γ is a Lagrange multiplier. By taking the partial derivatives and by equating them to zero, we obtain

$$\frac{\partial\lambda}{\partial\mu_i} = -\gamma\frac{\partial F}{\partial\mu_i}, \quad i=\overline{1, M}.$$

With the use of (7), (13), and (17), we obtain

$$\frac{1}{\mu_i}\frac{1}{\alpha_1}\lambda_1(N)[L_i(N)-L_i(N-1)]=-\gamma c_i a_i \mu_i^{a_i-1}, \quad i=\overline{1,M}. \tag{20}$$

For eliminating γ, we shall divide the i-th equation by the first. By virtue of (18) it then follows that

$$\frac{c_i a_i}{c_1 a_1}\frac{\mu_i^{a_i}}{\mu_a^{a_1}}=\frac{L_i(N)-L_i(N-1)}{L_1(N)-L_1(N-1)}, \quad i=\overline{2,M}, \tag{21}$$

$$\sum_{i=1}^{M} c_i \mu_i^{a_i}=S. \tag{22}$$

By substituting μ_i from (21) into (22), we finally obtain

$$\mu_1^{a_1}=S\Big/\Big\{c_1\Big\{1+\frac{1}{L_1(N)-L_1(N-1)}\sum_{i=2}^{M}\frac{a_1}{a_i}[L_i(N)-L_i(N-1)]\Big\}\Big\},$$

$$\mu_i^{a_i}=\frac{c_1 a_1}{c_i a_i}\mu_1^{a_1}\frac{L_i(N)-L_i(N-1)}{L_1(N)-L_1(N-1)}, \quad i=\overline{2,M}.$$

By using the formula

$$x_i\mu_i=x_1\mu_1\alpha_i/\alpha_1, \quad i=\overline{2,M}, \tag{23}$$

where x_1 can be set equal to $x_1 = 1$, we obtain the following equivalent system:

$$\mu_1^{a_1}=S\Big/\Big\{c_1\Big\{1+\frac{1}{L_1(N)-L_1(N-1)}\sum_{i=2}^{M}\frac{a_1}{a_i}[L_i(N)-L_i(N-1)]\Big\}\Big\}, \tag{24}$$

$$x_i^{a_i}=D_i\mu_1^{a_i-a_i}\frac{L_1(N)-L_1(N-1)}{L_i(N)-L_i(N-1)}, \quad i=\overline{2,M}, \tag{25}$$

where

$$D_i=\frac{c_i' a_i}{c_1 a_1}, \quad c_i'=c_i\left(\frac{\alpha_i}{\alpha_1}\right)^{a_i}. \tag{26}$$

It is often clearer to obtain the solution in terms of the relative occupation factors of the nodes. We can go over to the parameters μ_i ($i = \overline{2, M}$) with the aid of (23).

5. Minimization of Network Cost with a Capacity not Below the Assigned Value

It is required to minimize the objective function $F(\bar{\mu})$ under a capacity constraint $\lambda(N, \bar{\mu}) \geq \Lambda$. Just as in the case of maximization of the network capacity, we shall solve this problem by the method of undetermined Lagrange multipliers. Let us construct the Lagrange function $Q = F(\bar{\mu}) + \gamma[\lambda(N, \bar{\mu}) - \Lambda]$. We shall take the partial derivatives and equate them to zero: $\partial F/\partial\mu_i = -\gamma\partial\lambda/\partial\mu_i$, $i = \overline{1, M}$.

With the use of (7), (13), and (19), we obtain

$$C_i a_i \mu_i^{a_i-1}=-\gamma\frac{1}{\alpha_1}\frac{1}{\mu_i}\lambda_1(N)[L_i(N)-L_i(N-1)], \quad i=\overline{1,M}.$$

For eliminating γ, we shall divide the i-th equation by the first equation. By virtue of the formula $\lambda = \mu_1 U_1(N)/\alpha_1$ and of the constraint (19) we hence obtain

$$\mu_1=\Lambda\alpha_1/U_1(N),$$

$$\mu_i^{a_i}=\frac{c_1 a_i}{c_i a_i}\mu_1^{a_1}\frac{L_i(N)-L_i(N-1)}{L_1(N)-L_1(N-1)}, \quad i=\overline{2,M}.$$

By using formula (23), and by setting $x_1 = 1$, we can go over to the following equivalent system of equations:

$$\mu_1=\Lambda\alpha_1/U_1(N), \tag{27}$$

$$x_i^{a_i} = D_i \mu_1^{a_i - a_i} \frac{L_1(N) - L_1(N-1)}{L_i(N) - L_i(N-1)}, \quad i = \overline{2, M}, \tag{28}$$

where D_i is specified by (26).

The program of solution of the systems of nonlinear equations (24)-(25) and (27)-(28) written in the language PL/1 uses a subroutine of minimization of a function with M variables. An approximate solution can be obtained with a guaranteed error with respect to the functional.

6. Optimization of Capacity of Multiprogram Computers

Figure 1 shows a model of a computer system (CS) in the form of a closed network with M nodes, where the first node (the central processor CP) is the center of service, whereas the other nodes ($i = \overline{2, M}$) simulate the operation of $M - 1$ external devices ED. The arrival of an assignment at the first node (CP) corresponds to the execution of a sequence of instructions between two successive accesses of the external devices. The probability of calling the i-th ED is equal to P_i ($i = \overline{2, M}$); the quantity $P_0 = 1 - \Sigma P_i$ ($i = \overline{2, M}$) is the probability of completion of an assignment. In this case, a new assignment will arrive in the system via the branch denoted in Fig. 1 by "feeding a new assignment." Thus the number of calls (assignments) in the network remains constant, being equal to the multiprogramming level n.

It is assumed that all the assignments carried out by the system belong to the same class, and that the users' memory is shared out equally among n statistically identical assignments. Since the assignments require equal amounts of memory, it follows that the capacity of the main memory will be a linear function of n. The average time of service of a call in the i-th node is equal to $1/\mu_i$, $i = \overline{1, M}$, and the service rate can depend on the local queue length, i.e., we are given the quantities $\mu_i(n)$, where $n = \overline{1, N}$.

In the following the cost of a CS will signify the cost of its computational center which includes the processor, as well as the main memory and the external memory. The total cost S' of the CS and the cost S of the computational center are connected by the formula $S' = S + S_{per}$, where S_{per} is the cost of the peripheral equipment of a CS which includes the input-output devices with punched data carriers such as printers, videoterminals, data transmission multiplexers, and other CS devices needed for organizing the computational process.

The original parameters used for the designing are as follows: M is the number of nodes; N is the number of calls (assignments); the P_i are the transition probabilities, $i = \overline{1, M}$; V_1 is the average number of instructions carried out by the processor in each utilization; V_i is the average number of words transmitted in an input-output operation in the i-th external device, $i = \overline{2, M}$; K_i is the cost coefficient in the node i, $i = \overline{1, M}$; a_i is the nonlinearity factor in the node i, $i = \overline{1, M}$; S_p is the cost of the main memory share; S is the constraint on the CS cost (used in solving the design problem in the formulation 1); Λ is a constraint on the CS capacity (used in solving the design problem in the formulation 2).

A criterion of optimality in designing according to Formulation 1 is to maximize the CS capacity measured by the number of assignments carried out per unit time, under a constraint on the CS cost. As we noted, this cost does not include the cost of auxiliary equipment and installations, or operating expenditure.

We shall define the CS cost as follows:

$$F = \sum_{i=1}^{M} K_i b_i^{a_i} = \sum_{i=1}^{M} c_i \mu_i^{a_i} + M(n),$$

where $c_i = K_i(\omega_i/\alpha_i)^{a_i}$, K_i and c_i being cost coefficients in the node i; b_i is the speed of the server in the i-th node; μ_i is the rate of service in the i-th node; the α_i are the relative average rates of calling the nodes; ω_i is the total number of calls for processing needed per assignment in a node; $M(n) = nS_p$ is the cost of the main memory as a function of the multiprogramming level.

The average number of calls for processing per access to the i-th node is expressed by the formula $V_i = \omega_i/\alpha_i$, where $i = \overline{2, M}$; $\alpha_1 = 1/P_1$; $\alpha_i = P_i/P_1$, $i = \overline{2, M}$.

The relationship between the service rate μ_i and the server speed b_i is

$$\mu_i = \frac{\alpha_i}{\omega_i} b_i, \quad i = \overline{1, M}. \tag{29}$$

TABLE 1. Parameters of Work Load and Cost Estimate of CS Devices

Device	Transition probabilities, P_i	No. of operations in millions (transmitted words in millions) per access, V_i	Cost coeff., K_i	Exponent, a_i
CP	0,05	0,020	1 187 870	0,52716
ED	0,50	0,001	778 700	0,64572
ED	0,30	0,001	778 700	645 719
ED	0,15	0,001	2 021 690	0,96651

TABLE 2. Optimum Capacity of CS with the Original Load Parameters Listed in Table 1 and a CS Cost Constraint S = 1 Million Rubles

Multiprogramming level	Optimal capacity, sec^{-1}	Av. response time, sec	Processor speed, mill. oper. per sec	Speed of external devices, mill. words per sec		
				ED 2	ED 3	ED 4
2	0,8239	2,427	0,3959	0,03448	0,0247	0,0154
3	0,8744	0,3912	0,0271	0,0187	0,0187	0,0118
4	0,8693	4,600	0,3760	0,0222	0,0150	0,0095
5	0,7320	6,830	0,3177	0,0142	0,0116	0,0059
6	0,7926	7,569	0,3321	0,0159	0,010	0,0066
7	0,7398	9,462	0,3074	0,0138	0,009	0,0056
8	0,6836	11,703	0,2823	0,0120	0,008	0,0048

TABLE 3. Parameters of Work Load and Cost Estimate of CS Devices

Device	Transition probabilities, P_i	No. of operations in millions (transmitted words in millions) per access, V_i	Cost coeff., K_i	Exponent, a_i
CP	0,02	0,020	118 770	0,52716
ED	0,052	0,001	778 700	0,64572
ED	0,31	0,001	778 700	0,64572
ED	0,15	0,001	778 700	0,64572

TABLE 4. Optimum Capacity of CS with the Original Load Parameters Listed in Table 3 and a CS Cost Constraint S = 1 Million Rubles

Multi-Programming level	Optimal capacity, sec^{-1}	Av. response time, sec	Processor speed, mill. oper. per sec	Speed of external devices, mill. words per sec		
				ED 2	ED 3	ED 4
2	0,3167	6,314	0,3807	0,0345	0,0244	0,0150
3	0,3360	8,928	0,3768	0,0270	0,0186	0,0107
4	0,3340	11,976	0,3621	0,0221	0,0149	0,0087
5	0,3221	15,522	0,3426	0,0187	0,0124	0,0071
6	0,3048	19,686	0,3202	0,0160	0,0105	0,0059
7	0,2846	24,592	0,2965	0,0138	0,0090	0,0050
8	0,2631	30,408	0,2720	0,0120	0,0078	0,0043

The problem of optimizing the CS can be expressed in one of the following formulations.

Formulation 1. Find

$$\max \lambda = \mu_i P_i G(n-1)/G(n)$$

under the constraints

$$F = \sum_{i=1}^{M} c_i \mu_i^{a_i} + M(n) \leqslant S, \quad \bar{\mu} > 0.$$

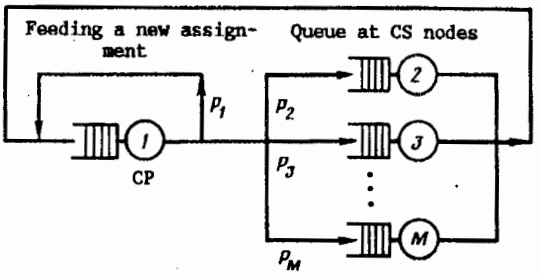

Fig. 1

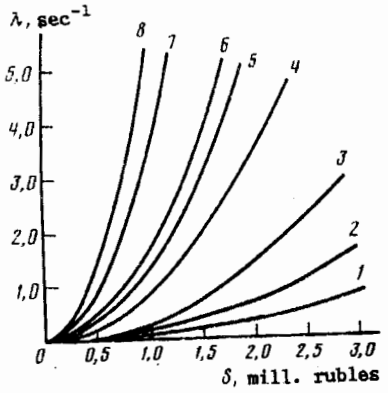

Fig. 2

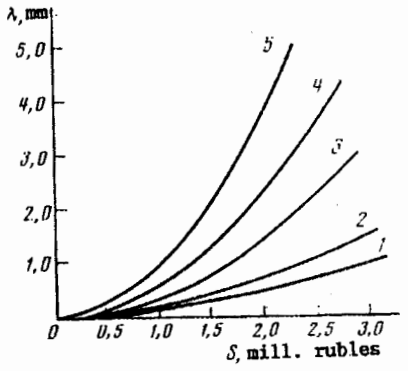

Fig. 3

Fig. 1. Model with central server.

Fig. 2. Optimum capacity plotted versus cost of computer system for various values of P_1. 1) $V_1 = 0.02$ million operations, $P_1 = 0.005$, $\lambda = 0.0846\ S^{1.9774}$; 2) $V_1 = 0.02$ million operations, $P_1 = 0.01$, $\lambda = 0.1680\ S^{2.0388}$; 3) $V_1 = 0.02$ million operations, $P_1 = 0.02$, $\lambda = 0.3364\ S^{2.0624}$; 4) $V_1 = 0.02$ million operations, $P_1 = 0.05$, $\lambda = 0.8517\ S^{1.9973}$; 5) $V_1 = 0.02$ million operations, $P_1 = 0.08$, $\lambda = 1.3697\ S^{2.0041}$; 6) $V_1 = 0.02$ million operations, $P_1 = 0.1$, $\lambda = 1.7114\ S^{2.0412}$; 7) $V_1 = 0.02$ million operations, $P_1 = 0.2$, $\lambda = 3.5037\ S^{2.0352}$; 8) $V_1 = 0.02$ million operations, $P_1 = 0.3$, $\lambda = 3.892\ S^{2.0482}$.

Fig. 3. Optimum capacity plotted vs. cost of computer system for various values. 1) $V_1 = 0.08$, $P_1 = 0.02$, $\lambda = 0.1081\ S^{2.0332}$; 2) $V_1 = 0.005$, $P_1 = 0.02$, $\lambda = 0.1613\ S^{2.0346}$; 3) $V_1 = 0.02$, $P_1 = 0.02$, $\lambda = 0.3364\ S^{2.0642}$; 4) $V_1 = 0.01$, $P_1 = 0.02$, $\lambda = 0.5676\ S^{2.0139}$; 5) $V_1 = 0.005$, $P_1 = 0.02$, $\lambda = 0.9051\ S^{2.0741}$.

Formulation 2. Find

$$\min F = \sum_{i=1}^{M} c_i \mu_i^{\alpha_i} + M(n)$$

under the constraints

$$\lambda = \mu_i P_i G(n-1)/G(n) \geqslant \lambda, \quad \mu > 0,$$

where g(n) is a normalizing constant of a closed network.

The solution is sought on the set of values of the service rates μ_i ($i = \overline{1, M}$) and of the multiprogramming levels n.

The vector of access $\bar{\alpha} = (\alpha_1, \ldots, \alpha_M)$ and the vector of calls for processing of assignments $\bar{\omega} = (\omega_1, \ldots, \omega_M)$ specify the parameters of the work load. The multiprogramming level n and the speed vector $\bar{b} = (b_1, \ldots, b_M)$ of the servers constitute (M + 1) variables. Let us note that after solving the optimization problem with respect to μ, we go over to the parameters b with the aid of (29).

Since the multiprogramming level n is usually confined to fairly small positive numbers, it is possible to select the optimal n in the optimization problem by discrete search. By taking n as an input parameter, we can solve the optimization problem with respect to the other M variables. The optimal solution will be the one (out of n feasible solutions) that has the best objective function.

For the model represented in Fig. 1, we listed in Tables 1 and 3 the parameters of the work load and estimates of the cost of the devices. The quantity V_1 is given in millions of operations, whereas V_i is given in millions of words, $i = \overline{2, M}$. The cost coefficients and the exponents have been obtained by regression analysis of models of standard computers.

In Table 1 it is assumed that the nodes 2 and 3 are supplemented with magnetic disk memories (MDM) with a capacity of 100 Mbyte, whereas the node 4 is supplemented with an MDM of 29 Mbyte capacity. In Table 3 it is assumed that the nodes 2 and 4 are supplemented with memories of 100 Mbyte.

The speed of the input–output devices is expressed in millions of words transmitted per second (with allowance for the time needed for fixing the heads, the search for the data on a track, and the data transmission; it corresponded to a unit of 1000 word capacity).

In Tables 2 and 4 we listed the results of optimization of the CS capacity with a cost constraint S = 1 million rubles, for the original load parameters listed in Tables 1 and 3, respectively. The cost of the main memory share was S_p = 37.5 thousand rubles.

The solutions (24)-(25) were sought from an initial point $x_i^0 = 1$, $i = \overline{2, M}$. The initial approximation for μ_1^0 was specified by (24). After obtaining the solution vector (μ_1, x_2, ..., x_M) the values of μ_i for $i = \overline{2, M}$ were obtained from (23), whereas the speed b_i ($i = \overline{1, M}$) of the devices was obtained from (16). The average response time was expressed by the formula

$$T = [N - L_1(N)]/\lambda_1(N).$$

In Tables 2 and 3 the optimum capacity corresponds to a multiprogramming level n = 3.

In Figs. 2 and 3 we plotted the optimum capacity vs. the CS cost for various values of the load parameters.

The curves were obtained by a regression analysis of the plots of the optimum capacity vs. the CS cost. The form of the curves shows that this relationship is close to the Grosch law.

APPENDIX 1

Proof of Theorem 1. The normalizing constant of a closed network with N calls and M nodes can be calculated [11], by the formula

$$G(N) = \sum_{n=0}^{N} f_i(n)\, g_M^i(N-n), \tag{A.1}$$

where

$$f_i(n) = e_i^n \Big/ \prod_{k=1}^{n} \mu_i(k) = x_i^n \Big/ \prod_{k=1}^{n} \beta_i(k);$$
$$g_M^i = \sum_{\substack{\bar{n} \in S(N, M) \\ n_i = N-n}} \prod_{\substack{j=1 \\ j \neq i}} f_j(n_j), \tag{A.2}$$

$\beta_i(k) = \mu_i(k)/\mu_i$, $k = \overline{1, N}$, $\mu_i(k)$ being the rate of service in the i-th node in the case of k calls in the queue; the set S(N, M) of all the feasible states of a system for a closed network with N calls and M nodes can be expressed as follows:

$$S(N,M)=\left\{ (n_1,\ldots,n_i,\ldots,n_M) \Big/ \sum_{i=1}^{M} n_i=N,\ n\geqslant 0,\ i=\overline{1,M} \right\}.$$

The term $g_M^i(n)$ is the normalizing constant of a network obtained from the original network by eliminating the i-th node, and only with n calls. Let us note that $g_M^M(n) = g_{M-1}(n)$ for $n = \overline{0, N}$.

By using (A.1)-(A.2) and (4), we obtain

$$\frac{\partial G(N)}{\partial x_i}=\partial\left(\sum_{n=0}^{N} f_i(n)\,g_M^i(N-n)\right)\Big/\partial x_i=\frac{1}{x_i}\sum_{n=1}^{N} n f_i(n)\,g_M^i(N-n)= \tag{A.3}$$

$$=\frac{1}{x_i}G(N)\sum_{n=1}^{N} nP_i(n)=\frac{1}{x_i}G(N)L_i(N),$$

where

$$P_i(n)=\frac{1}{G(N)}f_i(n)\,g_M^i(N-n),\qquad n=\overline{0,N}. \tag{A.4}$$

By virtue of (1) and (A.3) we can prove (10):

$$\frac{\partial\lambda_i(N)}{\partial x_i}=e_i\partial\left(\frac{G(N-1)}{G(N)}\right)\Big/\partial x_i=[e_i/G^2(N)]\left[G(N)\frac{\partial G(N-1)}{\partial x_i}-G(N-1)\frac{\partial G(N)}{\partial x_i}\right]=$$

$$=\frac{e_i}{G^2(N)}\left[G(N)\frac{1}{x_i}G(N-1)L_i(N-1)-G(N-1)\frac{1}{x_i}G(N)L_i(N)\right]=-\frac{1}{x_i}\lambda_i(N)[L_i(N)-L_i(N-1)].$$

By virtue of (2) and (A.3) we can similarly prove (8).

For proving (9) we obtain, with the use of (A.2)-(A.4), the formula

$$\frac{\partial P_i(n)}{\partial x_i}=\partial\left(\frac{1}{G(N)}f_i(n)\,g_M^i(N-n)\right)\Big/\partial x_i= \tag{A.5}$$

$$=\frac{1}{G^2(N)}\left\{G(N)\partial[f_i(n)\,g_M^i(N-n)]/\partial x_i-f_i(n)\,g_M^i(N-n)\frac{\partial G(N)}{\partial x_i}\right\}=$$

$$=\frac{1}{G^2(N)}\left\{G(N)n\frac{1}{x_i}f_i(n)\,g_M^i(N-n)-f_i(n)\,g_M^i(N-n)\frac{1}{x_i}G(N)L_i(N)\right\}=\frac{1}{x_i}[nP_i(n)-L_i(N)P_i(n)].$$

By using (4) and (A.5), we can prove (9):

$$\frac{\partial L_i(N)}{\partial x_i}=\partial\left[\sum_{n=1}^{N} nP_i(n)\right]\Big/\partial x_i=\frac{1}{x_i}\sum_{n=1}^{N}[n^2P_i(n)-L_i(N)nP_i(n)]=\frac{1}{x_i}\sum_{n=1}^{N}[n-L_i(N)]^2P_i(n)=\frac{1}{x_i}D_i(N).$$

Formulas (5)-(7) can be proved in the same way, either by direct differentiation, or by virtue of (8)-(10) and (3). For example,

$$\frac{\partial\lambda_i(N)}{\partial\mu_i}=\frac{\partial\lambda_i(N)}{\partial x_i}\frac{\partial x_i}{\partial\mu_i}=-\frac{\lambda_i(N)}{x_i}[L_i(N)-L_i(N-1)]\left(-\frac{x_i}{\mu_i}\right)'=\frac{1}{x_i}\lambda_i(N)[L_i(N)-L_i(N-1)].$$

The monotonicity of the functions follows from the fact that $D_i(N) \geq 0$ and $L_i(N) \geq L_i(N-1)$ for $\mu_i(k+1) \geq \mu_i(k)$, $i = \overline{1, M}$, $k = \overline{1, N}$.

APPENDIX 2

Proof of Theorem 2. For proving (16), we shall use (1) and (A.3):

$$\frac{\partial\lambda_i(N)}{\partial x_j}=\partial(e_iG(N-1)/G(N))/\partial x_j=$$

$$=\frac{e_i}{G^2(N)}\left[G(N)\frac{1}{x_j}G(N-1)L_j(N-1)-G(N-1)\frac{1}{x_j}G(N)L_j(N)\right]=-\frac{1}{x_j}\lambda_i(N)[L_j(N)-L_j(N-1)].$$

From (2) and (A.3) we similarly obtain (14)

$$\frac{\partial U_i(N)}{\partial x_j} = \partial(x_i G(N-1)/G(N))/\partial x_j =$$

$$= \frac{x_i}{G^2(N)}\left[G(N)\frac{1}{x_j}G(N-1)L_j(N-1) - G(N-1)\frac{1}{x_j}G(N)L_j(N)\right] = -\frac{1}{x_j}U_i[L_j(N)-L_j(N-1)].$$

For proving (15), we shall use (A.2)-(A.4) and hence obtain

$$\frac{\partial P_i(n)}{\partial x_j} = \partial\left(\frac{1}{G(N)}f_i(n)g_M{}^i(N-n)\right)\Big/\partial x_j = \frac{1}{G^2(N)}\left\{G(N)\partial(f_i(n)g_M{}^i(N-n))/\partial x_j - f_i(n)g_M{}^i(N-n)\frac{\partial G(N)}{\partial x_j}\right\} = \qquad (A.6)$$

$$= \frac{1}{G^2(N)}\left\{ G(N)f_i(n)\frac{1}{x_j}L_j{}^i(N-n)g_M{}^i(N-n) - f_i(n)g_M{}^i(N-n)\frac{1}{x_j}L_j(N)G(N)\right\} = \frac{1}{x_j}\{L_j{}^i(N-n)P_i(n) - L_j(N)P_i(n)\},$$

where $L_j{}^i(n)$ is the average queue length in the j-th node of a network with n calls that differs from the original network by the elimination of the node i; i, j = $\overline{1, M}$, n = $\overline{1, N}$.

By using (4) and (A.6), we obtain

$$\frac{\partial L_i(N)}{\partial x_j} = \partial\left(\sum_{n=1}^{N} nP_i(n)\right)\partial x_j = -\frac{1}{x_j}\left\{L_i(N)L_j(N) - \sum_{n=1}^{N}L_j{}^i(N-n)nP(n)\right\}.$$

For a network with two nodes we have $\frac{\partial L_i(N)}{\partial x_j} = -\frac{1}{x_j}D_i(N)$, $i \neq j$.

By using the formula $\sum_{i=1}^{M} L_i(N) = N$ and (9), we have $\sum_{\substack{i=1\\ i\neq j}}^{M}\frac{\partial L_i(N)}{\partial x_j} = \frac{1}{x_j}D_j(N)$. There hence follows the formula (15) and the monotonicity of $L_i(N)$.

Formulas (11)-(13) can be proved on the basis of (14)-(16) and (3) in the same way as in the proof of Theorem 1.

LITERATURE CITED

1. G. P. Basharin and A. L. Tolmachev, "Theory of queuing networks and its use in the analysis of data-processing systems," in: Itogi Nauki i Tekhniki [in Russian], Vol. 21, VINITI, Moscow (1983), pp. 1-119.
2. V. M. Vishnevskii, "Theory of queuing networks and its use in the analysis and design of computer systems and networks," in: Multiprocessor Computer Systems [in Russian], Inst. Control Problems, Moscow (1985), pp. 16-19.
3. V. M. Vishnevskii and A. S. Tverdokhlebov, "Models of closed networks with blocking used in the analysis of computer systems," Avtomat. Telemekh., No. 5, 172-179 (1980).
4. V. A. Zhozhikashvili and V. M. Vishnevskii, "The SIRENA computer network: planning and analysis," in: 10th All-Union Seminar School on Computer Networks [in Russian], Science Council on the Complex Problem "Cybernetics," Academy of Sciences of the USSR, Moscow (1985), pp. 377-382.
5. V. A. Zhozhikashvili, V. M. Vishnevskii, and M. G. Vinarskii, The Buffer Store of Switching Nodes in Computer Networks (Analysis and Design Methods) [in Russian], Preprint, Inst. Control Problems, Moscow (1986).
6. Z. L. Kruglyi, "Iteration algorithms of design of computer system configurations," in: Development and Utilization of Control Systems in Enterprises of the Radio-Engineering, Electronic, Instrument Construction, and Machine Building Industries [in Russian], Abstracts of Reports of Sci. Techn. Conference, Part 2, Machine Building Institute, Mogilev (1981), pp. 29-30.
7. Z. L. Kruglyi, "An algorithm of design of models of computer system configurations with various classes of assignments," Upravlyayushchie Sistemy Mashiny, No. 4, 73-79 (1980).
8. V. N. Kaminskii, "Optimization of closed stochastic networks with exponential service," Izv. Akad. Nauk SSSR, Tekh. Kibern., No. 6, 68-76 (1980).

9. Yu. I Mitrofanov, V. G. Belyakov, and V. Kh. Kurmangulov, Methods and Software of Analytic Simulation of Network Systems [in Russian], Science Council on the Complex Problem "Cybernetics," Academy of Sciences of the USSR, Moscow (1982).
10. É. Ya. Peterson and T. L. Plotkina, "Symmetry and entropy in optimizing the distribution of files," Avtomatika Vychisl. Tekhnika, No. 6, 7-13 (1983).
11. S. C. Bruell and G. Balbo, Computational Algorithm for Closed Queueing Network, North-Holland, New York (1980).
12. K. S. Trivedi and R. E. Kinicki, "A model for computer configuration design," Computer, 47-54 (1980).

METHOD OF RECONSTRUCTING THE STATE VECTOR OF A NONLINEAR DYNAMICAL SYSTEM FROM THE RESULTS OF OBSERVATIONS

V. I. Kushnarev and L. N. Lysenko

UDC 62-501.5:519.71

A method for solving the problem of reconstructing the complete state vector of dynamical systems described by ordinary nonlinear differential equations is given. In the method the scheme of solution of a two-point boundary problem of definite type is realized and the generalized algorithm for conditionally optimal filtration is used.

1. Introduction. Formulation of the Problem

In this paper we give one possible approach to the solution of the problem of reconstructing the complete state vector of a system whose dynamics are described by ordinary nonlinear differential equations satisfying existence and uniqueness conditions for solutions on a fixed interval of time.

We consider a nonlinear dynamical system, whose mathematical model, in general, has the following form:

$$\dot{\mathbf{x}}(t)=\mathbf{f}(t,\mathbf{x})+l\xi(t), \tag{1}$$

$$\mathbf{z}(t)=\varphi(t,\mathbf{y})+\eta(t). \tag{2}$$

Here $\mathbf{x}=\{\lambda,\mathbf{y}\}^{\mathrm{T}}$ is the n-dimensional state vector of the system; $\mathbf{z}$ is the m-dimensional vector of measurements; λ is a k-dimensional and $\mathbf{y}$ is an l-dimensional state subvector.

We shall have in mind that $n=l+k$, $n>m$, $\mathbf{x}\in\mathbf{X}^n$, $\mathbf{z}\in\mathbf{Z}^m$, $\lambda\in\Lambda^k\in\mathbf{X}^n$, $\mathbf{y}\in\mathbf{Y}^l\in\mathbf{X}^n$, where $\mathbf{X}^n$, $\mathbf{Z}^m$, Λ^k, $\mathbf{Y}^l$ are closed nonempty sets.

The vectors $\eta(t)$ and $\xi(t)$ on the right sides of (1) and (2) are n- and m-dimensional stochastic processes, whose necessary statistical characteristics are given.

The functions f and φ are continuous and continuously differentiable in all of their arguments on a fixed time interval.

It is assumed that the m-dimensional vectors corresponding to (2) are a certain collection of measurable parameters, functionally connected with the subvector y of the current state vector.

It is required, from the results of direct measurements of $\mathbf{z}(t)$ to estimate the state vector $\mathbf{x}(t)$ of the nonlinear dynamical system (1) on a fixed time interval, the components of the subvector $\lambda(t)$ of which are not connected with $\mathbf{z}(t)$ by finite analytic relations.

Moscow. Translated from Avtomatika i Telemekhanika, No. 2, pp. 54-61, February, 1987. Original article submitted February 17, 1986.

0005-1179/87/4802-0183$12.50