Scientific Research of the Institute of Mathematics and Computer Science, 2(10) 2011, 155-167

SOLUTION OF OPTIMAL CONTROL PROBLEM FOR THE THREE-LEVEL HM-NETWORK - II

***Mikhail Matalytski*[1], *Olga Kiturko*[2]**

[1]*Institute of Mathematics, Czestochowa University of Technology, Poland*
[2]*Faculty of Mathematics and Computer Science, Grodno State University, Belarus*
m.matalytski@gmail.com, sytaya_om@mail.ru

Abstract. A three-level HM queueing network with one type of requests and incomes, which is a stochastic model for goods transport in a logistics transport system is studied in this article. We studied the problems of control choice with and without a reduction of the current time in the case of a finite and infinite control horizon maximizing the expected income of a central system. We have compared three methods to find optimal control: the method of complete enumeration strategies, Bellman's method of dynamic programming and Howard's method.

Introduction

In article [1] presented in this journal, the investigation carried out of a closed Markov HM (Howard-Matalytski)-network with the same type of requests consisting of $M = n + m_1 + ... + m_{n-1}$ queueing systems (QS) S_i, $i = \overline{1,..n,1_1,...,1_{m_1},...,(n-1)_1, ...,(n-1)_{m_{(n-1)}}}$, which is a model of certain goods transportation is considered. In this model, central system S_n is the «producer» of a certain product; systems $S_1, S_2, \ldots, S_{n-1}$ are the «warehouses» where the product is stored; $S_{i_1}, S_{i_2}, \ldots, S_{i_{m_i}}$ are the «shops» (places of goods sale), which come from warehouse S_i, $i = \overline{1,(n-1)}$. The application here is seen as the shipment of goods in the logistics system (LS) «producer - warehouses - shops».

The system of equations for the expected incomes of the central system is obtained in a matrix form:

$$V_n(k,t+\Delta t) = \hat{Q}_n(k,t,\Delta t) + \hat{A}_n(k,\Delta t)V_n(k,t) \tag{1}$$

where $V_n(k,t)$ is a column vector of expected incomes for central system S_n during time t, if initially the network was in state k consisting of components $v_n(k,t)$ written down at the network states; $\hat{A}_n(k,\Delta t) = \left\|\hat{a}_{ij}(k,\Delta t)\right\|_{L\times L}$ is the matrix of transition probabilities between the states of the network over time Δt, if the initial network state was (k,t); and $\hat{Q}_n(k,t,\Delta t)$ is the column vector of the average

single-step income received by system S_n during time Δt, if at time moment t the network state was (k,t). Matrix $\hat{A}_n$ and vector $\hat{Q}_n$ can be found using matrix P, the intensities of requests service in QS, and incomes from the transitions between network states [1].

In paper [1] an analogue of (1) was obtained taking into account the fact that the amount in S c.u. (conventional units) being received during time Δt is equivalent to βS c.u. at the present time, the amount in S c.u. being received during n years is equivalent to $\beta^n S$ c.u. at the present moment. Coefficient $\beta \in (0,1]$ is called the re-evaluation (reduction) of future income coefficient. The analogue of system (1) in a matrix form is written as:

$$V_{n,\beta}(k,t+\Delta t) = \hat{Q}_n(k,\Delta t) + \beta \hat{A}_n(k,\Delta t) V_{n,\beta}(k,t) \tag{2}$$

In paper [1], the optimal control problem is formulated as well. Let us denote by θ_l for the control strategy in state l m, and let us consider $\Theta_l = \{\theta_l\}$ - the set of strategies in state l, $l = 1,2,...,L$. The vector of strategies $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, ..., \bar{\theta}_L) \in \Theta_1 \times \Theta_2 \times ... \times \Theta_L$ is called policy, where $\bar{\theta}_l$ is the chosen strategy in state l. If strategy θ_l or policy $\bar{\theta}$ are selected at time t, then we write $\theta_l(t)$ or $\bar{\theta}(t) = (\bar{\theta}_1(t), \bar{\theta}_2(t), ..., \bar{\theta}_L(t))$. The sequence of selected policies at every time moment forms control $\bar{\bar{\theta}} = (\bar{\theta}(t), \bar{\theta}(t+\Delta t), ..., \bar{\theta}(T_{\max}))$. If $T_{\max} < \infty$, it describes the finite control horizon, otherwise - the infinite one.

Let us consider that $E = E(\bar{\bar{\theta}})$ is the network functioning efficiency at a given control interval. Then the control $\bar{\bar{\theta}}^*$ that maximizes efficiency is called the optimal one. The optimal control problem for an HM network is to find optimal control:

$$E(\bar{\bar{\theta}}^*) = \max_{\bar{\bar{\theta}}} E(\bar{\bar{\theta}}) \tag{3}$$

As $E(\bar{\bar{\theta}})$, we can take the central system S_n income found using relations (1), (2). In [1], problem (3) is solved by the total exhaustive method of control strategies.

1. Application of Bellman's dynamic programming method

Let us consider the case where in every network state we can apply u control strategies $\theta_{l1}, \theta_{l2}, ..., \theta_{lu}$, for the sake of simplicity, we denote them $\theta_1, \theta_2, ..., \theta_u$. Let $P(\theta_s) = \left\| p_{ij}(\theta_s) \right\|_{M \times M}$ be a matrix of requests transitions probabilities between the QS network using strategy θ_s; $R(\theta_s) = \left\| r_{ij}(\theta_s) \right\|_{M \times M}$ is a matrix of one-step

income, $r_{ij}(\theta_s)$ is the income of system S_i, respectively, it is as well the waste or loss of system S_j when using strategy θ_s; $r(\theta_s)=\|r_i(\theta_s)\|_{1\times M}$ is the vector of constant incomes by using strategy θ_s, $r_i(\theta_s)$ is the income of system S_i per time unit if the network remains in the same state as when using θ_s; $\mu_i(\theta_s)$ is the intensity of requests for service in system S_i if we use strategy θ_s; $\mu(\theta_s)=(\mu_1(\theta_s),\mu_2(\theta_s),...,\mu_n(\theta_s))$, $s=\overline{1,u}$.

To determine the optimal control and corresponding to it the optimal expected income, we use the method of dynamic programming. In regards to equation (2), for a controlled network, we obtain:

– if $t_1=t+\Delta t$ then

$$\hat{Q}_n(k,\Delta t,\theta(t_1))+\beta\hat{A}_n(k,\Delta t,\theta(t_1))V_{n,\beta}(k,t)=V_{n,\beta}(k,t_1,\theta(t_1)) \tag{4}$$

where the matrix of transitions probabilities between the network states $\hat{A}_n(k,\Delta t,\theta(t_1))$ and the vector of average one-step incomes $\hat{Q}_n(k,\Delta t,\theta(t_1))$ determine further network functioning;

– if $t_2=t_1+\Delta t=t+2\Delta t$, then

$$\hat{Q}_n(k,\Delta t,\theta(t_2))+\beta\hat{A}_n(k,\Delta t,\theta(t_2))V_{n,\beta}(k,t_1,\theta(t_1))=V_{n,\beta}(k,t_2,\bar{\theta}(t_1),\theta(t_2)) \tag{5}$$

where $\bar{\theta}(t_1)$ from (5) is the strategy chosen at time t_1 from (4), and strategy $\theta(t_2)$ from (5) can be different at time t_2;

– if $t_3=t_2+\Delta t$, then

$$\hat{Q}_n(k,\Delta t,\theta(t_3))+\beta\hat{A}_n(k,\Delta t,\theta(t_3))V_{n,\beta}(k,t_2,\bar{\theta}(t_1),\theta(t_2))=$$

$$=V_{n,\beta}(k,t_2,\bar{\theta}(t_1),\bar{\theta}(t_2),\theta(t_3)) \tag{6}$$

at arbitrary t_m

$$\hat{Q}_n(k,\Delta t,\theta(t_m))+\beta\hat{A}_n(k,\Delta t,\theta(t_m))V_{n,\beta}(k,t_{m-1},\bar{\theta}(t_1),...,\bar{\theta}(t_{m-2}),\theta(t_{m-1}))=$$

$$=V_{n,\beta}(k,t_m,\bar{\theta}(t_1),...,\bar{\theta}(t_{m-1}),\theta(t_m)) \tag{7}$$

Let us first introduce some new notations: $\theta^*(t_i)$ is the optimal strategy at time t_i, $\bar{\theta}^*(t_i)$ is the optimal policy at time t_i, $V^*_{n,\beta}(t_i)$ is the optimal value of in the central QS expected income at time moment t_i. Then:

– if $t_1 = t + \Delta t$, considering (4) we have

$$V^{*}_{n,\beta}(k,t_1) = \max_{\theta(t_1)} V_{n,\beta}(k,t_1,\theta(t_1)) =$$

$$= \max_{\theta(t_1)}\left[\hat{Q}_n(k,\Delta t,\theta(t_1)) + \beta\hat{A}_n(k,\Delta t,\theta(t_1))V^{*}_{n,\beta}(k,t_0)\right] =$$

$$= \hat{Q}_n(k,\Delta t,\theta^{*}(t_1)) + \beta\hat{A}_n(k,\Delta t,\theta^{*}(t_1))V^{*}_{n,\beta}(k,t_0) \qquad (8)$$

where $V^{*}_{n,\beta}(k,t_0) = V_{n,\beta}(k,t_0)$ are defined, and thus we have $\theta^{*}(t_1)$ and $V^{*}_{n,\beta}(t_1)$;

– if $t_2 = t_1 + \Delta t = t + 2\Delta t$, considering (5) we have

$$V^{*}_{n,\beta}(k,t_2) = \max_{\bar{\theta}(t_1),\theta(t_2)} V_{n,\beta}(k,t_2,\bar{\theta}(t_1),\theta(t_2)) =$$

$$= \max_{\theta(t_2)} \max_{\bar{\theta}(t_1)}\left[\hat{Q}_n(k,\Delta t,\theta(t_2)) + \beta\hat{A}_n(k,\Delta t,\theta(t_2))V_{n,\beta}(k,t_1,\theta(t_1))\right] =$$

$$= \max_{\theta(t_2)}\left[\hat{Q}_n(k,\Delta t,\theta(t_2)) + \beta\hat{A}_n(k,\Delta t,\theta(t_2)) \max_{\theta(t_1)} V_{n,\beta}(k,t_1,\theta(t_1))\right] =$$

$$= \max_{\theta(t_2)}\left[\hat{Q}_n(k,\Delta t,\theta(t_2)) + \beta\hat{A}_n(k,\Delta t,\theta(t_2))V^{*}_{n,\beta}(k,t_1)\right] =$$

$$= \hat{Q}_n(k,\Delta t,\theta^{*}(t_2)) + \beta\hat{A}_n(k,\Delta t,\theta^{*}(t_2))V^{*}_{n,\beta}(k,t_1) \qquad (9)$$

and as a result we have $\theta^{*}(t_2)$ and $V^{*}_{n,\beta}(k,t_2)$;

– if $t_3 = t_2 + \Delta t$, considering (6) we have

$$V^{*}_{n,\beta}(k,t_3) = \max_{\bar{\theta}(t_1),\bar{\theta}(t_2),\theta(t_3)} V_{n,\beta}(k,t_3,\bar{\theta}(t_1),\bar{\theta}(t_2),\theta(t_3)) =$$

$$= \max_{\theta(t_3)} \max_{\bar{\theta}(t_1),\bar{\theta}(t_2)}\left[\hat{Q}_n(k,\Delta t,\theta(t_3)) + \beta\hat{A}_n(k,\Delta t,\theta(t_3))V_{n,\beta}(k,t_2,\bar{\theta}(t_1),\theta(t_2))\right] =$$

$$= \max_{\theta(t_3)}\left[\hat{Q}_n(k,\Delta t,\theta(t_3)) + \beta\hat{A}_n(k,\Delta t,\theta(t_3)) \max_{\bar{\theta}(t_1),\theta(t_2)} V_{n,\beta}(k,t_2,\bar{\theta}(t_1),\theta(t_2))\right] =$$

$$= \max_{\theta(t_3)}\left[\hat{Q}_n(k,\Delta t,\theta(t_3)) + \beta\hat{A}_n(k,\Delta t,\theta(t_3))V^{*}_{n,\beta}(k,t_2)\right] =$$

$$= \hat{Q}_n(k,\Delta t,\theta^{*}(t_3)) + \beta\hat{A}_n(k,\Delta t,\theta^{*}(t_3))V^{*}_{n,\beta}(k,t_2)$$

as a result we have $\theta^{*}(t_3)$ and $V^{*}_{n,\beta}(k,t_3)$.

Continuing this process for a number of steps 4, 5,..., $m-1$, we get optimal policies $\theta^{*}(t_1)$, $\theta^{*}(t_2),...,\theta^{*}(t_{m-1})$ and optimal values of the expected income $V^{*}_{n,\beta}(t_1)$, $V^{*}_{n,\beta}(t_2),...,V^{*}_{n,\beta}(t_{m-1})$.

Let us consider arbitrary time moment t_m ($t_m \le T_{\max}$). Considering (7) we have

$$V^*_{n,\beta}(k,t_m) = \max_{\bar{\theta}(t_1),\ldots,\bar{\theta}(t_{m-1}),\theta(t_m)} V_n(k,t_m,\bar{\theta}(t_1),\ldots,\bar{\theta}(t_{m-1}),\theta(t_m)) =$$

$$= \max_{\theta(t_m)} \max_{\bar{\theta}(t_1),\ldots,\bar{\theta}(t_{m-1})} \Big[\hat{Q}_n(k,\Delta t,\theta(t_m)) +$$

$$+ \beta\hat{A}_n(k,\Delta t,\theta(t_m))V_{n,\beta}(k,t_{m-1},\bar{\theta}(t_1),\ldots,\bar{\theta}(t_{m-2}),\theta(t_{m-1})\Big] = \max_{\theta(t_m)}\Big[\hat{Q}_n(k,\Delta t,\theta(t_m)) +$$

$$+ \beta\hat{A}_n(k,\Delta t,\theta(t_m)) \max_{\bar{\theta}(t_1),\ldots,\bar{\theta}(t_{m-2}),\theta(t_{m-1})} V_{n,\beta}(k,t_{m-1},\bar{\theta}(t_1),\ldots,\bar{\theta}(t_{m-2}),\theta(t_{m-1}))\Big] =$$

$$= \max_{\theta(t_m)}\Big[\hat{Q}_n(k,\Delta t,\theta(t_m)) + \beta\hat{A}_n(k,\Delta t,\theta(t_m))V^*_{n,\beta}(t_{m-1})\Big] =$$

$$= \hat{Q}_n(k,\Delta t,\theta^*(t_m)) + \beta\hat{A}_n(k,\Delta t,\theta^*(t_m))V^*_{n,\beta}(t_{m-1}) \qquad (10)$$

as a result we get $\theta^*(t_m)$ and $V^*_{n,\beta}(k,t_m)$. This process ends if $t = T_{\max}$, as a result we determine optimal control $\bar{\theta}^*$ and optimal income values $V^*_{n,\beta}(t)$.

The step by step algorithm for solving the problem by using Bellman's dynamic programming method with a finite control horizon looks as follows:

1) for each strategy θ_s, $s = \overline{1,u}$, using (8), we find the expression for $V_{n,\beta}(k,t_1,\theta(t_1))$ at moment t_1; out of all the founded expected incomes, we choose maximum expected income $V^*_{n,\beta}(t_1)$ and corresponding to it optimal strategy $\theta^*(t_1)$;
2) for each strategy θ_s, $s = \overline{1,u}$, using (9) and $V^*_{n,\beta}(t_1)$, we find expression $V_{n,\beta}(k,t_2,\theta(t_2))$ for moment t_2, out of all the found expected incomes we choose maximum expected income $V^*_{n,\beta}(t_2)$ and corresponding to it $\theta^*(t_2)$;
3) continuing the process at step $(m-1)$, we get optimal policies $\theta^*(t_1)$, $\theta^*(t_2)$,…, $\theta^*(t_{m-1})$, and the optimal values of expected incomes $V^*_{n,\beta}(t_1)$, $V^*_{n,\beta}(t_2)$,…, $V^*_{n,\beta}(t_{m-1})$. Using the obtained expected incomes and (10), we find $\theta^*(t_m)$ and $V^*_{n,\beta}(k,t_m)$;
4) this process ends if $t = T_{\max}$, as a result we determine optimal control $\bar{\theta}^*$ and the optimal value of income $V^*_{n,\beta}(T_{\max})$.

Example 1. Let us consider a network consisting of $M = 7$ QS with the number of requests $K = 11$, and let us assume that $\Delta t = 1$, $T_{\max} = 25$. The number of network states is $L = C^{7-1}_{7+11-1} = 12376$. Let us also assume that the «producer» is planning to conduct an advertising campaign. This may cause an increased number of goods shipments in the logistics system «producer - warehouses - shops», and

this may change other factors as well. Therefore, we consider two strategies: 1) to carry out an advertising campaign, and 2) not to carry out an advertising campaign.

Depending on when strategy is chosen, the following matrices and vectors which we believe to be specified are going to be different. The matrix of requests transitions probabilities between QS using each strategy is:

$$P(\theta_1)=\begin{pmatrix} 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0.35 & 0.65 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad P(\theta_2)=\begin{pmatrix} 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.65 & 0.35 \\ 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrixes of one-step incomes during the requests transition between the QS of the network are respectively:

$$R(\theta_1)=\begin{pmatrix} 0 & 0 & 0 & 5.6 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 6 \\ 5.3 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8.6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } R(\theta_2)=\begin{pmatrix} 0 & 0 & 0 & 6 & 2.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 6.1 \\ 3.6 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.5 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The vectors of QS incomes per time unit time depending on the strategy are:

$$r(\theta_1)=(7,\ 6.8,\ 4.6,\ 2,\ 7.8,\ 6,\ 5.6)$$

$$r(\theta_2)=(6.3,\ 7.8,\ 5,\ 4.2,\ 8,\ 7.2,\ 6)$$

The intensity of requests service are:

$$\mu(\theta_1)=(0.48,\ 0.5,\ 0.37,\ 0.45,\ 0.32,\ 0.29,\ 0.35)$$

$$\mu(\theta_2)=(0.6,\ 0.53,\ 0.46,\ 0.35,\ 0.43,\ 0.4,\ 0.55)$$

for example $\mu_1(\theta_1)=0.48$ is the intensity of requests service in system S_1 using strategy θ_1.

Using the Mathematica package, a computer program was worked out which allows one to find optimal strategies in each network state. Some results of calculations are presented in Table 1, where number 1 in the right-hand column of the Table means that one should conduct an advertising campaign at the appropriate interval, and number 2 - not to conduct an advertising campaign. The number of time intervals equals $\frac{T_{\max}}{\Delta t}=25$.

Table 1

Choosing a strategy for a system «producer» at every interval

Network state	Result of strategy choice at every interval
(8,1,1,1,0,0,0)	2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2
(8,1,1,0,1,0,0)	2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
(8,1,1,0,0,1,0)	2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
(8,1,1,0,0,0,1)	2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1
…	….
(5,1,2,2,1,0,0)	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
(5,1,2,2,0,1,0)	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
(5,1,2,2,0,0,1)	2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

From Table 1 it is clear, for example, that at initial state (5,1,2,2,1,0,0), an advertising campaign should be conducted.

2. Howard method application for solution of optimal control problem

2.1. Optimal control construction without considering revaluation

If we insert into formula (1) the asymptotic relation

$$V_n(k,t) = V_n(k) + tG_n(k)$$

we get

$$V_n(k) + (t+\Delta t)G_n(k) = \hat{Q}_n(k,t,\Delta t) + \hat{A}_n(k,\Delta t)(V_n(k) + tG_n(k))$$

where the right-hand side is equal to

$$\hat{Q}_n(k,t,\Delta t) + \hat{A}_n(k,\Delta t)V_n(k) + t\hat{A}_n(k,\Delta t)G_n(k)$$

Using this equality we will obtain a system of L Howard inhomogeneous equations for the $2L$ unknown $V_n(k) = \|v_{ni}(k)\|_{1\times L}$ and $G_n(k) = \|g_n(k)\|_{1\times L}$:

$$V_n(k) + \Delta t G_n(k) = \hat{Q}_n(k,\Delta t) + \hat{A}_n(k,\Delta t)V_n(k) \qquad (11)$$

Regarding renaming the states and introducing vectors $V_n = (v_{n1} \quad v_{n2} \quad \ldots \quad v_{nL})^T$, $G_n = (g_{n1} \quad g_{n2} \quad \ldots \quad g_{nL})^T$, $\hat{Q}_n(\Delta t)$ - the column-vector of average single-step incomes for the renamed states, and $\hat{A}_n(\Delta t)$ - matrix of transitions probabilities between the renamed states we will obtain in a matrix form

$$V_n + \Delta t G_n = \hat{Q}_n(\Delta t) + \hat{A}_n(\Delta t) V_n \tag{12}$$

The absolute values of weights v_{ni}, $i = \overline{1, L}$ from (12) cannot be determined, but it is possible to define the so-called relative weight, resting $v_{nj} = 0$, $j = j_1, j_2, \ldots, j_{L/2}$. Then we will obtain a system of L equations with L unknowns, which has a unique solution in the form of profit g_{ni}, $j \neq j_1, j_2, \ldots, j_{L/2}$ and relative weights v_{ni}, $j \neq j_1, j_2, \ldots, j_{L/2}$. It is important to stress that system (12) and its solution do not depend on t.

The economic meaning of relative weights can be easily understood from the form of asymptotic relations for the expected income. Let us take any two states i and j, for them

$$v_{ni}(t) = v_{ni} + t g_{ni} \text{ and } v_{nj}(t) = v_{nj} + t g_{nj}$$

hence

$$v_{ni}(t) - v_{nj}(t) = v_{ni} - v_{nj} = (v_{ni} + g_{ni}) - (v_{nj} + g_{nj})$$

i.e. the difference for the expected income generated by system S_n at initial network states i and j, for large t is the difference of the relative weights and profits. It shows how much more profitable it is to start exploitation of the network in state i than state j.

The Howard algorithm consists of two units - Estimation Control Unit (ECU), and Improvement Control Unit (ICU). In the first one, there are founded profits and relative weights for fixed control $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, ..., \bar{\theta}_L)$, that allow us to determine the average one-step income $q_{ni}^{(\bar{\theta}_i)} = \sum_{j=1}^{L} a_{ij}^{(\bar{\theta}_i)} r_{n,ij}^{(\bar{\theta}_i)}$, where $R_n(\bar{\theta}_s) = \left\| r_{nij}^{(\bar{\theta}_s)} \right\|_{L \times L}$ is a matrix of one-step incomes, $r_{nij}^{(\bar{\theta}_s)}$ is the income of system S_n if it changes state from i to j and used strategy $\bar{\theta}_s$. Then we can write Howard equation (11) as

$$V_n + \Delta t G_n = \hat{Q}_n^{(\bar{\theta}_i)}(\Delta t) + \hat{A}_n^{(\bar{\theta}_i)}(\Delta t) V_n \tag{13}$$

where quantities $\hat{Q}_n^{(\bar{\theta}_i)}(\Delta t)$ and $\hat{A}_n^{(\bar{\theta}_i)}(\Delta t)$ are assumed to be known in the ECU. The solutions of system (13) are values $v_{ni}(\bar{\theta})$ and $g_{ni}(\bar{\theta})$ that are uniquely appropriate to control $\bar{\theta}$. Profit values $g_{ni}(\bar{\theta})$ are the estimation of control $\bar{\theta}$ quality, explaining the name of the unit.

In the second unit, the ICU there finds the control ensuring higher profits with fixed weights. Let us consider the weights assigned arbitrarily (for example $v_{ni} = 0$ $\forall i$) or obtained in the ECU ($v_{ni} = v_{ni}(\bar{\theta})$). From system (12), for each i we have

$$G_n = \frac{1}{\Delta t}\left(\hat{Q}_n^{(\bar{\theta}_i)}(\Delta t) + \hat{A}_n^{(\bar{\theta}_i)}(\Delta t)V_n - V_n \right) \tag{14}$$

where quantities v_{ni} are assumed known for all i. Let us find a control $\bar{\theta}_i'$ that maximizes (14) for all $\bar{\theta}_i$, or equivalently, that maximizes the criterion

$$G_0 = \frac{1}{\Delta t}\left(\hat{Q}_n^{(\bar{\theta}_i)}(\Delta t) + \hat{A}_n^{(\bar{\theta}_i)}(\Delta t)V_n \right) \tag{15}$$

If you solve the problem of maximizing (15) for all $i = 1, 2, ..., L$, then you will obtain control $\bar{\theta}' = (\bar{\theta}_1', \bar{\theta}_2', ..., \bar{\theta}_L')$, which gives no less profit than control $\bar{\theta}$ with weights $v_{ni}(\bar{\theta})$.

A bunch of ECU and ICU, as amended by auxiliary units of choice control unit (CCU) $\bar{\theta}$, the choice of weights unit (CWU) v_{ni} and cycle organization unit (COU) forms the Howard iterative algorithm. The flow chart of this algorithm is shown in Figure 1.

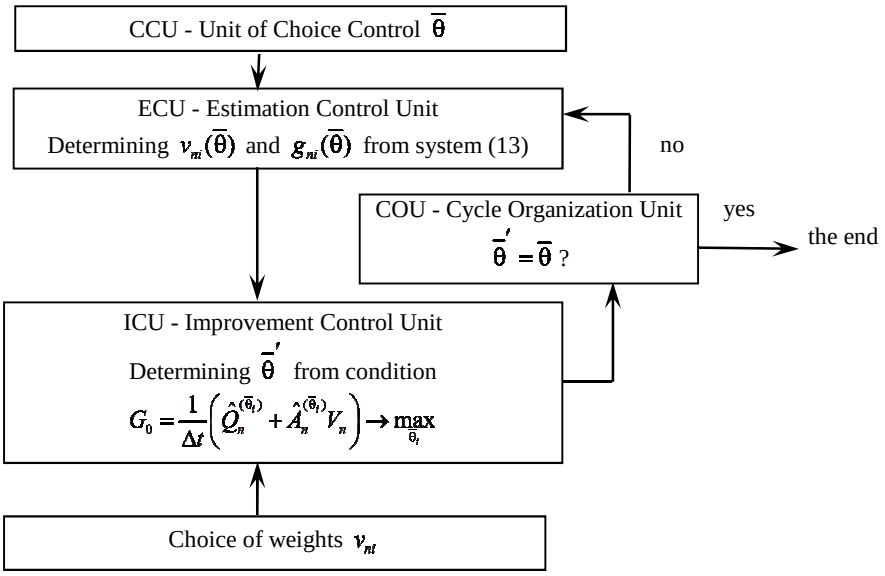

Fig. 1. Scheme of Howard iterative algorithm

The algorithm starts functioning with either a CCU or with a CWU. In the first case weights $v_{ni}(\bar{\theta})$ and profits $g_{ni}(\bar{\theta})$, $i = \overline{1, L}$ are found in the ECU. Using these data an attempt to improve control $\bar{\theta}$ in the ICU is assumed. If it succeeds, i.e. $\bar{\theta} \neq \bar{\theta}$, the ECU and ICU are cyclically finished though the CCU, otherwise the CCU stops the iteration process, and obtained value $\bar{\theta}$ together with $g(\bar{\theta})$ are announced to be optimal, i.e. $\bar{\theta}^* = \bar{\theta}$.

2.2. Optimal control construction with considering revaluation

Let us assume that $t \to \infty$, the optimal control is to be sought in the class of stationary. We take into account income revaluation $(\beta < 1)$, and as a criterion for optimal control, we take the limit income as defined in [1]:

$$V_{n,\beta,\infty}(\bar{\theta}) = \hat{Q}_n^{(\bar{\theta})}(\Delta t)\left(1 - \beta \hat{A}_n^{(\bar{\theta})}(\Delta t)\right)^{-1} \tag{16}$$

Considering revaluation, recurrence relation (2) allows us to pass to limit $t \to \infty$ and get

$$V_{n,\beta,\infty} = \hat{Q}_n(\Delta t) + \beta \hat{A}_n(\Delta t) V_{n,\beta,\infty} \tag{17}$$

We have the system of L Howard equations for L variables $v_{n,\beta,\infty,1}$, $v_{n,\beta,\infty,2}$, …, $v_{n,\beta,\infty,L}$. It underlies the Howard iterative algorithm of optimal control (optimal policy) construction. As in the previous section, the algorithm consists of two main units - ECU and ICU.

In the first one at fixed control $\bar{\theta}$, a system of equations is solved

$$V_{n,\beta,\infty} = \hat{Q}_n^{(\bar{\theta})}(\Delta t) + \beta \hat{A}_n^{(\bar{\theta})}(\Delta t) V_{n,\beta,\infty} \tag{18}$$

concerning marginal incomes $V_{n,\beta,\infty}$, however, $\hat{Q}_n^{\bar{\theta}}(\Delta t)$ and $\hat{A}_n^{\bar{\theta}}(\Delta t)$ are assumed to be known. Having the solution of the system, $V_{n,\beta,\infty}(\bar{\theta})$ will be obtained so that it unambiguously assesses control $\bar{\theta}$. Hence the name of the unit is the Estimation Control Unit.

In the second unit, the ICU concerned with the fixed values of limited income, for example, that are zero or obtained in ECU $(V_{n,\beta,\infty} = V_{n,\beta,\infty}(\bar{\theta}))$, for all i strategies $\bar{\theta}_i$ are defined so that maximize the criterion

$$V_0 = \frac{1}{\Delta t}\left(\hat{Q}_n^{(\bar{\theta}_i)}(\Delta t) + \hat{A}^{(\bar{\theta}_i)}(\Delta t) V_n\right) \to \max_{\bar{\theta}_i} \tag{19}$$

If we solve the problem of maximizing (19) for all $i = 1, 2, ..., L$, we will obtain control $\bar{\theta}' = (\bar{\theta}_1', \bar{\theta}_2', ..., \bar{\theta}_N')$, which gives an income limit not less than control $\bar{\theta}$, to which limit incomes $V_{n,\beta,\infty}(\bar{\theta})$ correspond.

As in the previous section, a bunch of ECU and ICU as amended by auxiliary units and a unit of the choice control unit (CCU), the choice of limit incomes unit (CLIU) and cycle organization unit (COU) forms the Howard iterative algorithm.

The algorithm starts working with either the CCU or CLIU. In the first case the limit incomes $v_{\beta,\infty,i}$ $\forall i$ are found in the ECU on the assumption that the selected in the CCU control $\bar{\theta}$, $\hat{Q}_n^{(\bar{\theta})}$ and $\hat{A}_n^{(\bar{\theta})}$ are formed.

Using the founded values of $v_{\beta,\infty,i}$ in the ICU we attempt to find control $\bar{\theta}'$ which maximizes criterion (19) for each $i=1,2,...,L$ and thereby improving control $\bar{\theta}$. If it succeeds, i.e. $\bar{\theta}' \neq \bar{\theta}$, the ECU and ICU cyclically finished through the COU, otherwise the COU stops the iterations, and result $\bar{\theta}$ obtained along with $v_{\beta,\infty,i}(\bar{\theta})$ is announced to be optimal, i.e. $\bar{\theta}^* = \bar{\theta}$.

If the algorithm starts from the CLIU (for example, the limit income for all i and control $\bar{\theta}$ are equal to zero), then in the ICU control $\bar{\theta}' \neq 0$ is determined and an iterative process is organized through the COU.

Example 2. Let us suppose that the system «producer» is planning to give a discount on freight. Let us consider a logistics system, consisting of $M=9$ QS and the number of requests is $K=8$, $\beta=0.55$, $\Delta t=1$. The number of states is $L=C_{9+8-1}^{9-1}=12870$. Therefore, we must consider two strategies: 1) to make a discount on freight, 2) not to make discounts on freight. However, this may change other factors as well. Let the vectors of systems incomes per time unit, depending on the strategy be:

$$r(\theta_1)=(3.4,\ 5,\ 1.9,\ 6.2,\ 4,\ 7,\ 3.4,\ 5.7,\ 6)$$

$$\text{and } r(\theta_2)=(4.6,\ 0.9,\ 2.5,\ 6.9,\ 0.5,\ 8,\ 4.6,\ 5.7,\ 6.9)$$

The matrix of requests transitions probabilities between network QS be as follows:

$$P(\theta_1)=\begin{pmatrix} 0 & 0 & 0 & 0.35 & 0.1 & 0.55 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.7 & 0.1 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P(\theta_2)=\begin{pmatrix} 0 & 0 & 0 & 0.35 & 0.3 & 0.35 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.4 & 0.35 \\ 0.45 & 0.55 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the matrixes of one-step incomes for the request transition between the QS respectively are:

$$R(\theta_1)=\begin{pmatrix} 3.4 & 0 & 0 & 6 & 8.2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 8.3 \\ 3 & 5.7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 5.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8.3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 7.2 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 4.8 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 5.9 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix} \text{ and}$$

$$R(\theta_2)=\begin{pmatrix} 4.6 & 0 & 0 & 6 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 3 \\ 3 & 7 & 2.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 6.9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 6.5 & 0 & 0 & 0 \\ 0 & 0 & 6.8 & 0 & 0 & 0 & 4.6 & 0 & 0 \\ 0 & 0 & 5.1 & 0 & 0 & 0 & 0 & 5.5 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 5.9 \end{pmatrix}$$

The intensity of requests service in the systems under these strategies are respectively the following:

$$\mu(\theta_1)=(0.7,\ 0.5,\ 0.41,\ 0.5,\ 0.02,\ 0.4,\ 0.73,\ 0.62,\ 0.31)$$

$$\text{and } \mu(\theta_2)=(0.6,\ 0.5,\ 0.36,\ 0.7,\ 0.41,\ 0.47,\ 0.56,\ 0.5,\ 0.19)$$

According to the Howard iterative algorithm taking into account revaluation, the following strategies ensuring at $T_{\max}\to\infty$ maximum limit income obtaining are optimal.

Table 2

Choosing strategy in LS depending on initial state of network

Network states	Result of strategy choice
(3,0,2,2,0,0,0,1,0)	1
(3,0,2,2,0,0,0,0,1)	2
(3,0,2,1,2,0,0,0,0)	2
(3,0,2,1,1,1,0,0,0)	2
(3,0,2,1,1,0,1,0,0)	2
(3,0,2,1,1,0,0,1,0)	1
(3,0,2,1,1,0,0,0,1)	2
…	…

Network states	Result of strategy choice
(0,0,1,1,0,0,2,4,0)	2
(0,0,1,1,0,0,2,3,1)	2
(0,0,1,1,0,0,2,2,2)	1
(0,0,1,1,0,0,2,1,3)	1
(0,0,1,1,0,0,2,0,4)	1
(0,0,1,1,0,0,1,5,0)	1
(0,0,1,1,0,0,1,4,1)	2
…	…

3. Comparison of three methods for solving optimal control problems

Considering the optimal control problem of an HM-network with incomes, three methods were used. They are: the method of complete enumeration of strategies [1], dynamic programming method and the Howard method. The applying conditions of these methods are shown in Table 3.

Our calculations for various HM-networks have shown that the method of complete enumeration for the solution of an optimal control problem has solved it almost three times faster than the Bellman method of dynamic programming. However, the method of dynamic programming allows one to more clearly show the choice of strategy at each intermediate interval, and using the method of complete enumeration, the strategy choice is carried out for the whole considered time interval.

Table 3

Applying conditions of methods for solving optimal control problem

Method of solution / Control Horizon	Method of complete enumeration		Bellman's dynamic programming method		Howard method	
	Considering revaluation	Without considering revaluation	Considering revaluation	Without considering revaluation	Considering revaluation	Without considering revaluation
Infinite	+	+	–	–	+	–
Finite	+	+	+	–	–	+

We have also discovered, that the method of complete enumeration is about one and a half times faster than the Howard method to solve the problem. However, with an increasing number of control strategies, the process of solving a problem by the complete enumeration method becomes more difficult, therefore it is better to obtain results by using the Howard method.

References

[1] Matalytski M., Kiturko O., Solution of optimal control problem for the three-level HM-network - I. Scientific Research of the Institute of Mathematics and Computer Sciences of Czestochowa University of Technology 2011, 1(10).