Optimization of Resource Allocation in the Distributed Production Network

Oleg Zaikin[1], Alexandre Dolgui[2], and Przemyslaw Korytkowski[1]

[1] Technical University of Szczecin,
Faculty of Computer Science and Information Systems,
Zolnierska 49, 71-210, Szczecin, Poland
ozaikine@wi.ps.pl
[2] University of Technology of Troyes,
12, rue Marie Curie, 10010, Troyes Cedex, France

Abstract. The development of the stochastic approach, based on queuing modelling and simulation, to solve the problem of resource decentralization in the distributed production network (DPN) is examined in the paper. The problem of performance optimisation of DPN is formulated as task of non-linear integer programming with stochastic parameters. A branch and bound algorithm, based on sequential use of queuing modeling and simulation, is designed for optimization of the complete model. The analytical modeling is used for low bound calculation and choice of optimal direction of branching. The simulation is used for verification of analytical results and definition of motion on variant tree.. Proposed approach is oriented towards corporate computer and telecommunication networks, High-Tec assembly manufacturing, printing and publishing industry.

1 Introduction

There are three main factors, which affect the development of new generation of computer aided manufacture (CAM) systems. They are:

1. Widespread penetration of information technologies into all levels of CAM systems. Information technologies (IT) are essentially changing output of global production. Last decade new kind of intelligent production has been developed as independent branch of industry. Peculiarity of intelligent production is, that its output is direct result of IT. They are High-Tec electronic production, print media industry, distribution of databases and application programs. Intelligent production have caused new methods of management and organization, making them highly distributed and networked.
2. Development of telecommunication media in CAM systems Distributed intelligent production can function well in developed telecommunication space It means firstly, that communication technology becomes fundamental part of production process and quality of communication servicing (QoS) mostly .affects the efficiency of intelligent production. Corporate network is important component of distributed network.

3. Production and consumption of intelligent product are customer-oriented and have address character. In other words intelligent production is the system of mass servicing of population. There can be many agents, participated in intelligent production, which differ geographically as well functionally. Behavioural models of agents are stiffly interrelated in production scenario and from this point of view intelligent production is multi-agent system. A typical task of distributed multi-agent system is resource allocation (productive, labour, material) in central office and peripheral nodes. This task appears each time, when the same functions can be perform in centre and on the spots. It can be distribution of authorities in organization structure, distribution of material flows in supply chain, distribution of multiplex channels in satellite telecommunication system.

2 Problem Statement

Lets examine a QN with star configuration. For example we have a big printing company including several regional (local) factories and a central one (Fig.1). Let us suppose that offset printing machines equip the local factories and more expensive digital printing machines equip a central one. As usual, the local factories serve customer demands. However when some of them is overcharged, they are transferred to centre. It is obviously, that the cost of central digital printing is more than cost of local offset one. On the other side the lead-time of central factory is considerably less than lead-time of local factory.

figure=Figure1.eps,height=8.6cm

Fig. 1. Graph representation of DQN

Let the rate of customer's demand for each local factory and processing time for all equipment are given. It is necessary to define the number of offset machines for each local factory and the number of digital machines for the central factory, minimising total cost of printing with restriction on customer's demand due date.

The studied system can be deemed as open queuing network (QN), which can be represented as an oriented graph. Vertices of the graph are processing nodes and arcs of the graph are flow processes between them.

In the studied case, we examine QN with star configuration, i.e. there are a set of local processing nodes (PN), servicing corresponding local areas and one network centre (NC), servicing global region. Each local PN consist of a number of private servers, operating in parallel input (IP) and output (OUT).

If all servers of corresponding local PN are occupied, the incoming jobs go to the NC node with limited number of common (leased) servers, operating all flow processes with more high productivity. If all leased servers of NC are busy, the incoming job goes to the waiting queue of the corresponding PN. The jobs at PN and NC are served in FCFS discipline. There are no restrictions on the number of jobs waiting at the queue.

It is necessary to define the number of private servers at each local PN and leased servers at the NC, taking into account, that local and leased servers has different productivity and costs of the idle time. Below the formal model of resources assignment is represented:

Input data

$S = \{s\}$ – a set of processing nodes (PN),
$F = \{f_s\}$– a set of flow processes, incoming at PN,
$\Pi(f) = \{\Psi_f(n,t), \lambda_f\}$ – parameters of FP 'f_s', where $\Psi_f(n,t)$ – distribution law of arrival of customer demands, λ_f – the rate of arrival,
$\widetilde{\tau}_s, \widetilde{\tau}_c$ - average servicing time of a job at local PN or NC (including time of delivering),
γ_s, γ_c – costs of a server (acquisition, exploitation, amortisation) reduced to unit time for private or leased server correspondingly,

Control parameters

$N_s, s \in S$ – a number of private servers, assigned for each local PN $s \in S$,
N_c – a number of leased servers, assigned to the NC.

Criterion function

We examine three components of criterion function at formulated task.

1. The total servicing time of all flow processes for period of optimisation T_0,

$$T_\Sigma = \sum_f \lambda_f \widetilde{\tau}_s T_0 \tag{1}$$

2. Total costs of equipment, reduced to period T_0,

$$C_\Sigma = (N_c\gamma_c + \sum_s N_s\gamma_s)T_0 \tag{2}$$

3. Total costs of servers utilisation during period T_0,

$$U_\Sigma = (\frac{\widetilde{\lambda}_c}{N_c\mu_c}\widetilde{\delta}_c + \sum_s \frac{\widetilde{\lambda}_s}{N_s\mu_s}\widetilde{\delta}_s)T_0 \tag{3}$$

where $f \in F$ is the index of flow process,
$s \in S$ – the index of PN,
λ_f – the rate of arrival jobs for flow process f,
$\widetilde{\tau}_s$ – the average servicing time of a job at PN s,
μ_s – the rate of servicing of jobs at PN s,
N_s – a number of private servers, assigned for PN s,
N_c – the number of leased servers, assigned for NC,
$\widetilde{\lambda}_s, \widetilde{\lambda}_c$ – effective arrival rate to private or leased server s correspondingly,
γ_s, γ_c – reduced to time unit cost of private or leased servers ,
$\widetilde{\delta}_s, \widetilde{\delta}_c$ – reduced to time unit costs of idle time for private or leased servers,
T_0 is a period of optimisation.

The first component T_Σ defines the total quality of servicing (customer point of view), the second C_Σ and third U_Σ ones defines the total costs of installation and servicing relatively (manager point of view). These components of criterion function depend on control parameters N_s and N_c in different ways. Nevertheless, it is possible to define such values of control parameters, which provide trade-off, i.e. minimum of total criterion function.

$$CR = \alpha T_\Sigma + \beta U_\Sigma + C_\Sigma = min \tag{4}$$

Here α and β are waging coefficients.

3 Solution method

3.1 Combination of simulation and optimization approach

The objective of proposed approach is to combine the analytical methods and simulation approach for optimization of IPN. We develop a complex systems optimization approach, which consist of following steps:

- choice of basic network models and their analytical verification,
- incremental construction of complete model while using the basic models,
- use of simulation techniques for validation of the complete model,
- jointly use of simulation and optimization algorithms for step by step optimization of the model, obtained in previous steps.

A branch and bound algorithm may be used for global optimisation of the complete model. The proposed algorithm is based on sequential use of analytical calculation and simulation [2], [11]. The developed optimisation algorithm applies the main idea of Branch-and-Bound method. The analytical model [14] is used for primary evaluation and choice of optimal direction of branching. The simulation is used for validation of analytical results Branch-and-Bound method comprises three components:

1. Method of the alternative variants tree construction
2. Method of calculation for low bound
3. Way of the motion on variants tree

The optimisation algorithm tries to reallocate a part of the resource from NC node to local PN:

- Choice of the quantity of resource to reallocate from NC node to local PN.
- Analytical calculation for each local PN of the criterion value CR.
- Choice of PN for which the criterion value is minimal; reallocate the resource from NC to the PN.
- Simulation for the chosen variant.

If CR value is better then continue this branch; otherwise return to previous branch. Branching is started from an initial capacity location where all the resources are assigned to NC node.

3.2 Validation of analytical model through simulation

More possibilities for analysis of distributed queuing system and making of optimal decision are provided by the simulation model [1], [11], which has not any restrictions for dimension, kind of arrival pattern, discipline and time of servicing. The objectives of simulation consist in validation of proposed in (3.1) analytical model and verification of proposed in (3.2) branch and bound algorithm.

To compose the simulation model the program ARENA will be applied, which is the integrated environment for stochastic processes with simple and efficient programming language. Additional advantage of ARENA is advanced tools for data visualisation [8].

Simulation experiment was conducted for the following conditions:

1. QN, consisting of arbitrary number of processing nodes and a one network-centre. It corresponds a multi-channel queuing system with two kinds of servers (private and leased). The private and leased servers have different productivity and maintenance cost.
2. The observation period is *1000 tu*. It ensures the truth of simulation trial. Provided simulation experiments shown that simulation which lasts longer than *500 tu* is stationary. Warm-up interval is *200 tu*.
3. The-simulation model is realised for:
   - the Poisson law of jobs arriving and exponential distribution of servicing time
   - the Poisson law of jobs arriving and deterministic distribution of servicing time
   - the Poisson law of jobs arriving and the Erlang law of the 2nd degree for servicing time distribution.
4. Discipline of servicing is First come-First served. The waiting queue is infinite. Running of each waiting job is repeated with interval τ_R.

Structure of simulation model in terms of program ARENA is shown on the (Fig.2)

To find global minimum value of the criterion function we have to examine three-dimensional space (Fig. 3). The first dimension is number of servers in the model, the second is their distribution among locals and central nodes, and finally the third dimension are values of criterion function.

Lets examine the network, which consists of one central and two local PN. The initial data are the following:

- input flows, entering each local PN are Poisson ones with arrival rates $\lambda_{S1} = 20$, and $\lambda_{S2} = 10$,
- leased and private servers are exponential, Deterministic or Ernang 2nd degree with rates of servicing $\mu_{S1} = \mu_{S2} = 10$, $\mu_c = 16$ and average service time $\widetilde{\tau}_{S1} = \widetilde{\tau}_{S2} = 0.1, \widetilde{\tau}_c = 0.016$ relatively,
- repeating time $\tau_R = 0.5$,
- waging coefficient $\alpha = 1$ and $\beta = 2$

figure=Figure2.eps,height=8.9cm

Fig. 2. Structure of production node

figure=Figure3.eps,height=7.4cm

Fig. 3. Search area for global minimum

Table 1 presents optimum configurations for particular type of provided analyses. In bold the optimal results for particular method of achieving are presented. Column Analytical presents searching optimum by using formula 1- 8. Column Exponential presents simulation results where servers have exponential servicing time. Column Deterministic presents simulation results where servers have fixed servicing time. Column Erlang 2nd order presents simulation results with the Erlang law of the 2nd degree for servicing time distribution. Column presents the traffic intensity in the network
Represented simulation examples have shown the following results:

1. Presented simulation examples proved adequacy of proposed analytical model. Conducted simulation experiments have shown that obtained by simulation network configuration is the same or very close to the optimal configuration calculated with analytical model. The degree of accuracy depends on the traffic intensity (ρ). The lower traffic intensity is, the closer results we obtain.
2. Changing type of server doesnt affect on the trend of searching optimal solution. Simulation experiments have shown that presented analytical model for network with exponential servers is valid also for network where servers are deterministic or Erlang 2nd degree. Comparing results for network with different kind of servers it follows that the results differ no more than 3

4 Conclusion

1. The problem of resources allocation in QN can be formulated as task of discrete optimisation of parameters in the queuing system. The objective function of the task depends on the stochastic and deterministic parameters, such as distribution law of arrival jobs in flow process, distribution law of service time, capacity of waiting queue, number of parallel servers in the queuing systems.
2. With some admissions about kind of flow process and service time an analytical solution can be obtained for primary evaluation of possible decision. Such analytical result is obtained in the article for multi-channel queuing system with different kinds of servers and unlimited capacity of waiting queue.
3. More possibilities for analysis of queuing system and making of optimal decision are provided by the simulation approach, which has not any restrictions for dimension, kind of jobs arrival, discipline and time of servicing. Conducted simulation experiment confirmed that value of objective function is critical to resource allocation in QN.

Table 1. Analytical and simulations result of optimal server distribution for model with various number of servers with one central node and two local nodes

NC	PN_1	PN_2	Analytical	Simulation			
				Exponential	Deterministic	Erlang 2nd	ρ
			5 servers				
2	2	1	16.2861	17.1439	16.7153	16.8894	0.4839
1	3	1	15.3351	17.1744	16.2719	16.7207	0.5357
			6 servers				
1	4	1	16.1345	16.9748	16.3897	16.6765	0.4546
1	3	2	16.2975	16.8121	16.3903	16.6125	0.4546
			7 servers				
1	5	1	17.0795	17.6608	17.2205	17.434	0.3947
1	4	2	17.0968	17.2618	17.0806	17.1856	0.3947
			8 servers				
1	5	2	18.0416	18.1247	18.0192	18.0551	0.3488

The simulation provided the following results:

- chosen criterion function are critical to server distribution between processing nodes and network-centre ,
- validation - simulation confirms the adequacy of proposed analytical methods,
- proposed analytical method may be used for evaluation and searching of most reasonable configuration of QN,
- simulation modelling has to be used for more accurate definition of the criteria value.

References

[1] Azadivar, F. and Lee, Y.: Optimization of discrete variable stochastic systems by computer simulation. Mathematics and Computers in Simulation, 2, (1988), 331-345

[2] Biethan, J. and Nissen, V.: Combinations of simulation and evolutionary algorithms in management science and economics. Annals of Operations Research, 32, (1994), 183-208

[3] Buzacott, J. and Shanthikumar, J.: Modeling and analysis of manufacturing systems. John Wiley & Sons, New York, (1993)

[4] Chee Hock Ng: Queuing Modelling Fundamentals. John Wiley & Sons, New York, (1997)

[5] Dolgui, A.: A Set of Models for Design and Study of Automated Production Systems. Ph.D. Dissertation, Institute of Engineering Cybernetics of Academy of Sciences of Belarus. (1990) (in Russian)

[6] Dolgui, A.: Utilisation de Y-transformation pour loptimisation des parametres des modeles de simulation. In: Actes de la Premiere Confrence Francophone de Modlisation et de Simulation (MOSIM97), Hermes, Paris, (1997), 451-459.

[7] Dolgui, A. and Ofitserov, D.: A Stochastic Method for Discrete and Continuous Optimization in Manufacturing Systems. Journal of Intelligent Manufacturing, 5, (1997), 405-413
[8] Gordon W., Newell G.: Closed Queuing Systems with Exponential Servers, Operation Research, v. 15, (1967), p.254-265
[9] Guariso, G., Hitz, M. and Werthner, H.: An integrated simulation and optimization modeling environment for decision support. Decision Support Systems, 1, (1996), 103-117
[10] Hall, R.W.: Queuing methods for service and manufacturing. Prentice Hall, Englewood Cliffs. N. Y, (1991)
[11] Kelton, W.D., Sadowski, R.P. and Sadowski, D.A.: Simulation with Arena. McGraw-Hill, N. Y., (1997)
[12] Zaikin, O. and Dolgui, A.: Resource assignment in mass demand HighTec assembly manufacturing based on the queuing modelling. In: Proc. International Conference on Industrial Logistics (ICIL'99), St. Petersburg, 1999, University of Southampton Publication, (1999), 200-209
[13] Zaikin, O., Dolgui, A. and Kraszewski P.: Simulation Approach to resource allocation in the satellite telecommunication network. In: Proc. 13th European Simulation Multi-conference, Warsaw, (1999), SCS Publication, Delft 1.
[14] Zaikin, O. and Ignatiev, V.: A method of analysis of multi-channel queuing models. Izv. AN SSSR (Technical Cybernetics, Academy of Science of USSR), 6, (1973), 86-88