On Local Optima Distribution in Buffer Allocation Problem for Production Line with Unreliable Machines ⋆⋆

Alexandre Dolgui [*] **Anton Eremeev** [**] **Viatcheslav Sigaev** [***]

[*] *IMT Atlantique, Nantes, France*
[**] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia (e-mail: eremeev@ofim.oscsbras.ru).*
[***] *Avtomatika-Servis LLC, Omsk, Russia (e-mail: sigvs@yandex.ru).*

Abstract: In this paper, we consider a manufacturing flow-line organized as a series-parallel system of machines separated by finite buffers. The failure and repair times of machines are supposed to be exponentially distributed. The production rate of each machine is deterministic, and different machines may have different production rates. The buffer allocation problem consists in determining the buffer capacities with respect to a given optimality criterion, which depends on the average production rate of the line, the buffer acquisition and installation cost and the inventory cost. The tentative solutions are evaluated with an approximate method based on the Markov models aggregation approach. The computational experiments show better quality of solutions obtained by a genetic algorithm compared with the local descent and tabu-search algorithms. It is indicated that in many test problems several clusters of local optima can be found.

Keywords: Production line, Unreliable machines, Buffer allocation, Series-parallel network, Genetic algorithms, Local optima.

1. INTRODUCTION

Buffer capacity allocation problems arise in a wide range of manufacturing systems, such as transfer lines, flexible manufacturing or robotic assembly systems which are flow lines. Buffers separate any two consecutive machines. The parts are accumulated in a buffer when the machines downstream are less productive than machines upstream. Assume that machines can breakdown. When a breakdown occurs, the corresponding machine is not used in production for a random repair time which is independent on the number of failed machines. It is assumed that there is a sufficient number of raw parts at the input of the system and these parts are always available. The completed parts depart from the system immediately. The performance of the flow-line is measured in terms of the average production rate, i.e., the steady state average number of parts produced per unit of time.

In the literature, there are two types of publications. The first concerns only evaluation of the line performance for a given size of buffers. In the second, the buffer sizes are optimized. For example, (Dallery and Gershwin, 1992), (Gershwin, 1993), (Heavey et al., 1993), (Meerkov and Li, 2008), and (Tan and Gershwin, 2009) proposed models to evaluate the performance of lines with unreliable machines and fixed sizes of buffers. Markov models and aggregation or decomposition techniques are often used to calculate steady state throughput or other performance indicators for these lines provided that the buffer capacities are given. Based on these models for performance analysis, in, e.g., (Smith and Daskalaki, 1988), (So, 1997), (Gershwin and Schor, 2000), and (Shi and Gershwin, 2009), the optimization for buffer capacity allocation was considered with respect to diverse optimality criteria and for different types of lines.

1.1 *The Buffer Allocation Problem Formulation*

In the present paper, we consider buffer allocation problem for line with a series-parallel network, as shown in Fig. 1, where the arrows are machines and the circles are buffers.

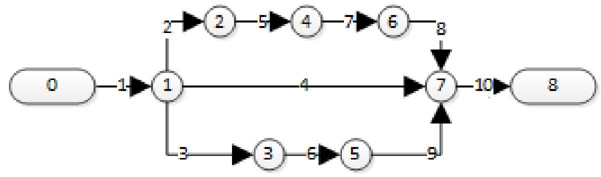

Fig. 1. Example of a line with a series-parallel network

We assume that a machine can be in an operational state or under repair. An operational machine may be blocked and temporarily stopped in case if there is no room in the downstream buffer. It may also be starved if there are no parts to process in the upstream buffer. Otherwise operational machines are working. In what follows, m denotes the number of machines in a production line. A working machine i, $i = 1, \ldots, m$, is assumed to have a

⋆ The research is supported by Russian Science Foundation grant 21-41-09017.
⋆⋆The research is supported by Russian Science Foundation grant 21-41-09017.

constant cycle time C_i and, then, the average production rate $u_i = 1/C_i$.

It is supposed that machines break down only when they are working. The times to fail and times to repair for each machine are assumed to be mutually independent and exponentially distributed random values. Let T_b^i denote the average time to fail, and let $\lambda_i = 1/T_b^i$ be the failure rate for working machine i, $i = 1, \ldots, m$. Similarly, let T_r^i and $\mu_i = 1/T_r^i$ denote respectively the time to repair and the repair rate for machine i. Under the above mentioned assumptions the system has the steady state mode (see e.g. (Sevast'yanov, 1962)), and performance of the system in this state is important for applications.

Let the buffers in the system be denoted by $B_1, \ldots, B_N$ and let h_j be the capacity of buffer B_j, which is to be decided. Denote the vector of decision variables as $H = (h_1, h_2, \ldots, h_N) \in Z_+^N$, where Z_+ is the set of non-negative integers.

Optimization criteria that was used is:

$$\max \phi(H) = T_{am} R(V(H)) - Q(H) - J(H), \qquad (1)$$

where

- T_{am} amortization time of the line (line life);
- $V(H)$ average production rate (steady state throughput);
- $R(V)$ revenue related to the production rate V;
- $J(H)$ cost of buffer configuration H;
- d_j maximal admissible capacity of buffer B_j, $j = 1, \ldots, N$;
- $Q(H) = c_1 q_1(H) + \ldots + c_N q_N(H)$ average steady state inventory cost, where $q_j(H)$ is the average steady state number of parts in buffer B_j, for $j = 1, \ldots, N$.

Function $\phi(H)$ has to be maximized, subject to the constraints $h_1 \leq d_1$, $h_2 \leq d_2$,..., $h_N \leq d_N$. $R(V)$ and $J(H)$ are assumed to be given monotone non-decreasing real-valued functions. The cost function $J(H)$ may be non-linear to model some standard buffer capacities or penalize solutions where the total capacity of the buffers exceeds an upper bound. A non-linear revenue function $R(V)$ can model the law of diminishing returns, for example, it can reflect the effect of overproduction by switching from strictly increasing to constant at a certain threshold. A stepwise revenue function can be used to model zero revenue in case of an unacceptably low average production rate (see e.g. Section 3).

The performance of this transfer line is measured in terms of production rate, i.e. the steady state average number of parts produced per unit time. For the evaluation of this parameter, different types of Markov models have been considered in literature (see e.g. Dallery and Gershwin, 1992; Papadopoulos and Heavey, 1996).

In general, the production rate with finite buffers is difficult to analyze precisely with the Markov models. Exact performance computation of a production rate of a line with more than two machines and one buffer is problematic due to exponential growth of the number of states. Therefore, most of the techniques employed for the analysis of such systems are in the form of analytic approximations and simulations. Analytical approximations are generally based on the two-machines-one-buffer Markov models, and either aggregation (De Koster, 1987) or decomposition approach (Dallery et al., 1989; Gershwin, 1987; Li, 2005). Simulation models are more expensive computationally but applicable to a wider class of systems (Dolgui and Svirin, 1995; Srensen and Janssens, 2004).

In this paper, we use two-machines-one-buffer Markov model independently developed by Levin & Pasjko (1969), Dubois & Forestier (1982), and Coillard & Proth (1984). For each tentative buffer allocation decision, the production rate is evaluated via an aggregation algorithm (Dolgui, 1993; Dolgui and Svirin, 1995), which is similar to the Terracol and David (1987) techniques. This aggregation approach appears to be sufficiently rapid for evaluation of tentative buffer allocations within the optimization algorithms.

The aggregation algorithm for production rate evaluation consists in recurrent replacement of two adjacent machines by a single machine. The parameters λ^*, μ^*, c^* of the resulting single machine are calculated from differential equations corresponding to the two-machines-one-buffer Markov model. After N steps of such aggregation procedure the system reduces to one virtual machine with parameters λ^*, μ^*, c^* and the estimate of the overall production rate $V(H)$ is given by $c^* \mu^/ (\lambda^* + \mu^)$

In our previous work, we proposed several metaheuristics (Dolgui et al., 2002; 2007) for some problems of this type. In these metaheuristics, we used a two-machine one-buffer Markov model (Levin & Pasjko, 1969; Dubois & Forestier, 1982; Coillard & Proth, 1984; and Dolgui, 1993) - see some elements necessary here in Appendix - and an aggregation algorithm (Dolgui, 1993; Dolgui and Svirin, 1995), which is similar to the Terracol and David (1987) techniques to evaluate the average production rate of each tentative buffer allocation decision for the more general case of series-parallel lines with more than two machines. This aggregation approach appears to be sufficiently rapid for the evaluation of tentative buffer allocations within the optimization algorithms.

1.2 Contribution of the Paper

Our main contribution consists in formulation of the packet routing problem as a fractional length-bounded maximum multicommodity flow problem where all edges have the unit length. Investigation of the practical performance of a simple greedy heuristic and of the FPTAS from **?** with respect to the secondary criterion and the required CPU time is another contribu- tion of the paper. Therefore it may be considered as a more detailed presentation and study of the practical instances considered in **?**.

2. INVESTIGATION OF LOCAL OPTIMA PROPERTIES

Experimental location studies local optima in the solution space of the optimization problem indicate that in many problems there is a high concentration of local optima in the immediate proximity to the global optimum **?**. This observation is known as the thesis about the existence of a “big valley” for minimization problems or “massive central” for tasks to the maximum Boese (1993); Hains et al. (2011). Distance to global optimum in our case, we will calculate in the metric l_1.

The formulation of this thesis can be divided into two parts:

(1) Local optima are located relatively close both to each other and to the global optimum.

(2) Values of the objective function of local optima tend to deteriorate with increasing distance to the global optimum.

This thesis partly explains the performance of genetic algorithms. If local optima are collected in the population and the next solution is chosen somewhere between two arbitrary local optima, then such a process has many chances to find global optimum. In this regard, verification and theoretical substantiation of this hypothesis is of undoubted interest.

In order to test this thesis, the paper proposes a method finding the number of integer points in the intersection ball of a given radius in the l_1 metric with a parallelepiped, having faces parallel to the coordinate planes. The the method was obtained by reducing the problem to a combinatorial one formulation already considered earlier for the case of non-negative integer points, using generating functions **?**. In the present work, a generalization the specified result to the conditions we need.

A series of experiments were carried out on existing tasks for determination of the structure of local optima. With a one-time running the local lifting algorithm LSA solution H whose elements h_i are chosen with uniform distribution between 1 and d_i improved to a local optimum. This procedure was repeated 300 times to create the necessary sample size **?**. Based on the results obtained, it was calculated the number of admissible solutions $\mid \Omega \mid$ in balls containing all found local optima. In this case, the distance from best found local optimum to the worst, and beyond the center - the best local optimum found.

In table 1-4 the column $V1$ contains the cardinality $\mid \Omega \mid$ for balls containing local optima, the column $V2$ contains the cardinality of the entire space solutions for each problem, and the column $V1/V2$ is their ratio. As can be seen from the tables, the first part of the thesis about the existence of a ¡¡central mountain range¿¿, that local optima are located relatively close both to each other and to the global optimum, in our case, it was not performed and mainly in the tasks used the optima are strongly scattered throughout the solution space.

№	$V1$	$V2$	$V1/V2$
1	2,84E+05	5,74E+05	0,49
2	8,88E+11	9,48E+11	0,94
3	8,60E+01	4,85E+03	0,02
4	2,34E+22	2,89E+22	0,81
5	4,17E+07	9,75E+07	0,43
6	1,00E+00	5,23E+08	0,00

Table 1. Results of running a local search multiple times for a series as.1 - as.6

№	$V1$	$V2$	$V1/V2$
9	7,97E+03	1,00E+04	0,80
10	7,55E+03	1,46E+04	0,52

Table 2. Results of running a local search multiple times for a series vp6.9 - vp6.10

№	$V1$	$V2$	$V1/V2$
9	6,50E+03	1,00E+04	0,65
10	9,95E+03	1,46E+04	0,68

Table 3. Results of running a local search multiple times for a series vp7.9 - vp7.10

Also considered was a part of the thesis about the tendency to worsen the value of the objective function of local optima with increasing distance to the global optimum. To do this, for all the considered problems, the correlation was determined $\rho(\varphi(\xi), r(\xi, \xi^*))$ objective function values local optima $\varphi(\xi)$ and the distance to the global optimum $r(\xi, \xi^*)$. As shown by experiments on the proposed problems, there is a negative correlation ρ, and all values correlations are significantly different from 0 with a confidence level of 95% (tables 5–6).

№	$V1$	$V2$	$V1/V2$
1	1,44E+05	1,94E+05	0,74
2	9,96E+03	1,94E+05	0,05
3	9,07E+04	1,94E+05	0,47
4	2,06E+04	1,94E+05	0,11
5	6,95E+04	1,94E+05	0,36
6	6,94E+04	1,94E+05	0,36
7	2,45E+04	1,94E+05	0,13
8	3,42E+03	1,94E+05	0,02
9	6,28E+04	1,94E+05	0,32
10	1,00E+00	1,94E+05	0,00

Table 4. Results of running a local search multiple times for a series bn5.1 - bn5.10

№	$\rho(r(\xi, \xi*), \varphi(\xi))$
Series as.1 - as.5	
1	-0,87078206
2	-0,437208613
7	-0,547348483
8	-0,943725905
Series vp6.9 - bn6.10	
9	-0,833583818
10	-0,778237651

Table 5. Results of running a local search multiple times for a series *as* and *vp*6

An interesting fact is that that in problems as.6, bn5.1 the entire set of local optima splits into clusters, for each of which a part of the thesis about the tendency to deterioration of the value of the objective function of local optima with increasing distance to the global optimum.

Figures 2 and 3 show diagrams of local optima for problems as.6, bn5.1, where the ordinate shows the value of the objective function of the local optimum $\varphi(H)$, and the abscissa is distance in metric l_1 to the chosen solution.

№	$\rho(r(\xi, \xi*), \varphi(\xi))$
Series bn5.1 - bn5.9	
1	-0,706451573
2	-0,872602935
3	-0,914939714
4	-0,999969666
5	-0,972555479
6	-0,729655068
7	-0,973827419
8	-0,846807033
9	-0,884916669
Series vp7.9 - bn7.10	
9	-0,907792301
10	-0,889724863

Table 6. Correlation results for series bn and $vp7$

The effect of clustering is especially well manifested in problems with parallel sections of the line. This effect can be justified by the fact that there are several different paths in lines with a parallel network. from start bin to end bin. Thus, if you create two lines that are identical in graph structure, such that one line will have relatively large bunkers along the same path, and the other - otherwise, it is possible to achieve two solutions close in value to the objective function, but located on ¡¡enough¿¿ great distance from each other.

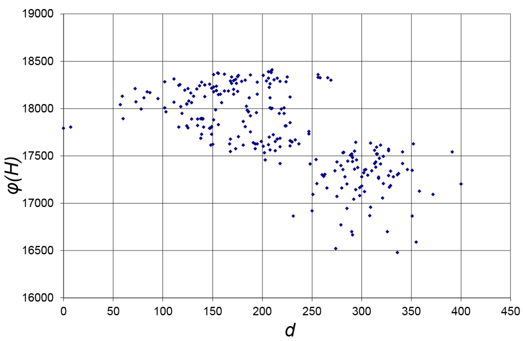

Fig. 2. The structure of local optima of the problem as6.

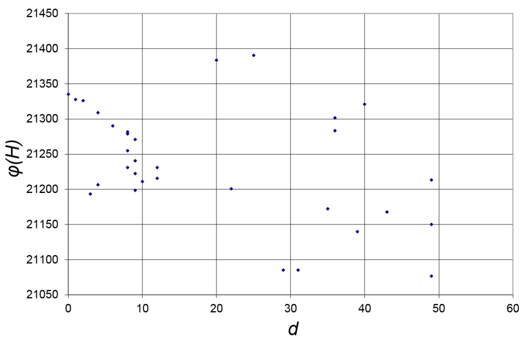

Fig. 3. The structure of local optima of the problem bn5.1.

Another reason for the clustering effect is the property line symmetry. This property is set to **?** for a serial line, consisting of two EU and a bunker between them. If we swap the parameters of the first and second EOs so that that the input section of the line will become the output, and the output - the input, then line performance will not change.

Consider the example of serial line t.1 from three HUs and two hoppers between them with the following parameters: $T_i^{\mathrm{O}} = T_i^{\mathrm{B}} = 1,\ i = 1, \ldots, 3;\ U_1 = 1,\ U_2 = 0.5\ ,\ U_3 = 1;$ $d_j = 4,\ c_j = 0,\ j = 1, \ldots, 2;\ T_{am} = 7000;\ J(H) = 50 \cdot (h_1 + h_2)$; if $V(H) < 2570$ then $R(V(H)) = 0.9 \cdot V(H)$, otherwise $R(V(H)) = 2570$.

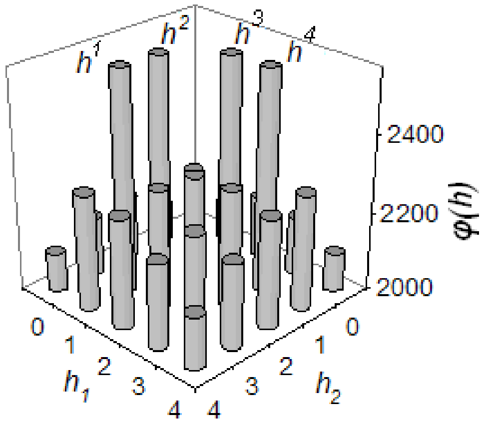

Fig. 4. Values of the objective function on the set D for problem t.1

The set of global optima (see Fig. 4) of this example splits into two clusters. IN the first cluster contains solutions $H^1 = (1, 2)$ and $H^2 = (1, 3)$, and the second is the solutions $H^3 = (2, 1)$ and $H^4 = (3, 1)$. Clustering optima in considered example is a consequence of the symmetry and singularities aggregation algorithm from **?** used in calculating the average line performance.

In the general case, for lines with $n \geq 3$ bins, the effect symmetry can be observed in several internal areas lines. This leads to the fact that with this aggregation algorithm the set of local optima is divided into several clusters.

3. RELAXED PACKET ROUTING PROBLEM

4. COMPUTATIONAL EXPERIMENT

For computational experiments, three series of problems were considered in the dissertation:

Series	Number of Lines	Number of HU	Source
AS	8	4 – 14	Renault Production **?**
BN	10	5	**?**
VP	4	5	**?**

Table 7. List of series

The AS series consists of tasks created from streaming lines 1,2,3,4,5,6,7,8 from **?** with real data from produced by Renault. Test case 3 from **?** was not used as it demonstrates a special case where approximate calculation of line performance is essential differs from the simulation results. Lines 4,5 from **?** are not used due to small dimensions. A distinctive feature of the 7.8 lines is the presence parallel sections (see figures **??**, **??**). Line parameters 1,2,6,7,8 are given in tables 8-**??**.

bn5.1-bn5.10 series consists of 10 lines with 5 machines in each. There are two ¡¡bottlenecks¿¿ in the problems of this series. Under ¡¡bottleneck¿¿ is understood as a line section of two relatively slow machines and a bunker between

Bunkers		EU			
i	d_i	j	T_j^{O}	T_j^{B}	U_i
1	20	1	244.2	150	10
2	17	2	255.3	300	10
3	38	3	176	75	10
4	48	4	184	600	10
		5	192	450	10

Table 8. Task parameters *AS1*

them. The parameters of all lines are given in the tables **??**-**??**.

Series vp6.9-vp6.10 and vp7.9-vp7.10 **?** are composed of problems on lines from 5 consecutive machines. Machine parameters and line structure are given in tables **??** - **??**.

4.1 Graph "Clique for links".

4.2 Graph "Large clique for links"

4.3 Graph"Star for Stations"

4.4 Graph "Clique for Stations"

5. EXPERIMENTAL COMPARISON OF ALGORITHMS FOR SOLVING THE PACKET ROUTING PROBLEM

6. CONCLUSION

(1) Approximate solution of the packet routing problem in software defined networks for many practical cases can be found using the proposed greedy algorithm with relatively low computing time. The computation time for most of the considered SDSN instances was less than a second.
(2) If the transfer rate of each session is negligibly small, compared to other problem input data, then the routing problem can be solved with any chosen accuracy using the FPTAS. The computation time for the considered practical instances was from tens of seconds to few thousand seconds.
(3) In practice the FPTAS obtains solutions with approximation ratio much smaller than its approximation guarantee.

REFERENCES

REFERENCES

Boese KD, Kahng AB and Muddu S (1993). On the big valley and adaptive multi-start for discrete global optimizations. Technical report, UCLA CS Department, TR-930015.

Hains, D., Whitley, L. & Howe, A. Revisiting the big valley search space structure in the TSP. J. Oper. Res. Soc. 62, 305-312 (2011)