3rd International Conference on Industrial Engineering and Industrial Management
XIII Congreso de Ingeniería de Organización
Barcelona-Terrassa, September 2nd-4th 2009

The Response Time Variability Problem: A Review

Albert Corominas[1], Alberto García-Villoria[1], Rafael Pastor[1]

[1] Institute of Industrial and Control Engineering (IOC), Universitat Politècnica de Catalunya, Av. Diagonal, 647, 08028. Barcelona, Spain. albert.corominas@upc.edu, alberto.garcia-villoria@upc.edu, rafael.pastor@upc.edu

Keywords: response time variability, regular sequences, scheduling, combinatorial optimisation

1. Introduction

The Response Time Variability Problem (RTVP) is a combinatorial optimisation problem that occurs whenever products, clients or jobs need to be sequenced so as to minimise variability in the time between the instants at which they receive the necessary resources. These situations can be generalized under the following framework. A sequence is built using n symbols (that represent products, clients, jobs, …) where symbol i (i = 1,...,n) is to occur given number d_i of times in the sequence (that represent the number of times that symbol i has to receive the resource). In the RTVP, the optimal solution is the sequence that minimises variability in the distances between any two consecutive copies of the same symbol. In other words, the distance between any two consecutive copies of the same symbol should be as regular as possible (ideally, constant).

A family of related problems in which would be included the RTVP can be formed by combining the following characteristics (León et al., 2003):

− Cyclic vs Non-cyclic. The problem is cyclic if the sequence is the same for all cycles and the distance, for each symbol i, between the first copy of i in a cycle and the last copy of i in the preceding cycle is considered.

− Distance-constrained vs Not distance-constrained. The problem is distance-constrained if the distance between two consecutive copies of the same symbol has an upper bound and/or a lower bound.

− Optimality vs. Feasibility. If the aim is to find a solution that optimises an objective function then we look for optimality. Instead, if the aim is to find a feasible solution, then we look for feasibility.

The RTVP is cyclic, not distance-constrained and its objective is to optimise an objective function. Its formulation is the following. Let n be the number of symbols, d_i the number of copies to be scheduled of symbol i (i = 1,…,n) and D the total number of copies (equal to $\sum_{i=1..n} d_i$). Let s be a solution of an instance in the RTVP that consists of a circular sequence of copies ($s = s_1 s_2 \ldots s_D$), where s_j is the copy sequenced in position j of sequence s. For each symbol i in which $d_i \geq 2$, let t_k^i be the distance between the positions in which the copies k + 1 and k of symbol i are found. We consider the distance between two consecutive positions to

be equal to 1. Since the sequence is circular, position 1 comes immediately after position D; therefore, $t_{d_i}^i$ is the distance between the first copy of symbol i in a cycle and the last copy of the same symbol in the preceding cycle. Let $\overline{t_i}$ be the desired average distance between two consecutive copies of symbol i ($\overline{t_i} = D/d_i$). Note that for each symbol i in which $d_i = 1$, t_1^i is equal to $\overline{t_i}$. The objective is to minimise the metric called response time variability (RTV), which is defined by the sum of the square errors with respect to the $\overline{t_i}$ distances. This is given in the following expression:

$$RTV = \sum_{i=1}^{n} \sum_{k=1}^{d_i} \left(t_k^i - \overline{t_i} \right)^2. \tag{1}$$

Note that the RTV metric is a weighted variance with weights equal to d_i. That is, $RTV = \sum_{i=1..n} d_i \cdot Var_i$, where $Var_i = 1/d_i \cdot \sum_{i=1..d_i} \left(t_k^i - \overline{t_i} \right)^2$.

An illustrative example is the following. Let $n = 3$ with symbols A, B and C. Also consider $d_A = 2$, $d_B = 2$ and $d_C = 4$; thus, $D = 8$, $\overline{t_A} = 4$, $\overline{t_B} = 4$ and $\overline{t_C} = 2$. Any sequence such that contains symbol i ($\forall i$) exactly d_i times is a feasible solution. For example, the sequence (C, A, C, B, C, B, A, C) is a feasible solution, and has an $RTV = \left((5-4)^2 + (3-4)^2 \right) + \left((2-4)^2 + (6-4)^2 \right) + \left((2-2)^2 + (2-2)^2 + (3-2)^2 + (1-2)^2 \right) = 12$.

The RTVP is difficult to be solved optimally; it has been proved to be NP-hard (Corominas et al., 2007).

The remainder of the paper is organized as follows: Section 2 identifies some real-world situations in which the RTVP occurs. Section 3 surveys the RTVP and summarises the main results obtained. Finally, some conclusions and suggestions for future research are exposed in Section 4.

2. Applications of the RTVP

This problem has a broad range of real-world applications. One of the first situations in which the idea of the regular sequence appeared was the sequencing of mixed-model assembly lines at Toyota Motor Corporation under the just-in-time (JIT) production system. Since Toyota popularized the just-in-time (JIT) production systems, the problem of sequencing on mixed-model assembly lines has acquired high relevance. One of the main aims of JIT is to eliminate sources of waste and inefficiency. In the case of Toyota, the main source of waste was the production of excessive volumes of stock. To solve this problem, JIT systems produce only the specific models required and in the quantities needed at any given time. According to Monden (1983), in this type of system the units should be scheduled in such a way that the consumption rates of the components in the production process remain constant. Miltenburg (1989) also studied this scheduling problem and considered only the demand rates for the models (Miltenburg, 1989; Kubiak, 1993). The problem proposed by Miltenburg intended to minimise variations in production rate in different models. However, feedback received from the manufacturing industry suggests that a good mixed-model sequence is one in which the distances between units of the same model are as regular as possible. One drawback of the Miltenburg problem is that, on the contrary of the RTVP, it takes the positions of the models with only one unit to be produced into account although the positions of these models are irrelevant for the regularity of the consumption rates.

The RTVP also appears in computer multithreaded systems (Waldspurger and Weihl, 1994 and 1995; Dong et al., 1998). Multithreaded systems (operating systems, network servers, media-based applications, etc.) do different tasks to attend to the requests of client programs that take place concurrently. These systems need to manage the scarce resources in order to service the requests of n clients. For example, multimedia systems must not display video frames too early or too late, because this would produce jagged motion perceptions (Kubiak, 2009). Waldspurger and Weihl, considering that resource rights could be represented by tickets and that each client i had a given number d_i of tickets, suggested the RTV metric to evaluate the sequence of resource rights.

Other contexts in which the RTVP can be applied are the design of sales catalogues (problem introduced in Bollapragada et al., 2004), the periodic machine maintenance problem (Anily et al., 1998; Wei and Liu, 1983) as well as other distance-constrained problems (e.g., see Han et al., 1996).

Two real-life cases of RTVP applications were reported in the literature. In Bollapragada et al. (2004), the study is motivated by the problem faced by the National Broadcasting Company (BNC) of U.S., one of the main firms in the television industry. Major advertisers buy to BNC hundreds of time slots to air commercials. The advertisers ask to BNC that the airings of their commercials are evenly spaced as much as possible over the broadcast season. The problem solved finally is not the RTVP, but a non-cycling variant. This study is continued in Brusco (2008). In Herrmann (2007), the author came up with the RTVP while working with a healthcare facility that needed to schedule the collection of waste from waste collection rooms throughout the building. Based on data about how often a waste collector had to visit each room and in view of the fact that different rooms require a different number of visits per shift, the facility manager wanted these visits to occur as regular as possible so that excessive waste would not collect in any room. For instance, if a room needed four visits per eight-hour shift, it should be ideally visited every two hours.

3. State of art

Although the RTVP is in general NP-hard, the two-symbol case can be optimally solved with a quick algorithm proposed in Corominas et al. (2007). For a general case, Corominas et al. (2007) proposed a mixed-integer linear programming (MILP) model whose practical limit to obtain optimal solutions is 25 copies to be sequenced. Corominas et al. (2009b) proposed an improved MILP model and increased the practical limit for obtaining optimal solutions from 25 to 40 copies to be sequenced. Thus, the use of heuristic and metaheuristic algorithms is justified.

This problem has been first time solved in Waldspurger and Weihl (1994) using a method that authors called lottery scheduling. This method is based on generating a solution at random as follows. For each position of the sequence, the symbol to be sequenced is chosen at random and the probability of each symbol is equal to the number of copies of this symbol that remain to be sequenced divided by the total number of copies that remain to be sequenced. The same authors proposed a greedy heuristic method that they called stride scheduling (Waldspurger and Weihl, 1995) that obtains better results than the lottery scheduling method. However, the stride scheduling method is, in fact, the Jefferson method originally designed to solve the apportionment problem (Balinski and Young, 1982; Bautista et al., 1996).

In Corominas et al. (2007) five heuristics are proposed to solve the RTVP: the bottleneck algorithm used in Moreno (2002) to solve the minmax Product Rate Variation problem, random generation, two classical parametric methods for solving the apportionment problem called Webster method and Jefferson method (Balinski and Shahidi, 1998) and a new heuristic called Insertion method by the authors; moreover, a local search procedure is applied

to the solutions obtained with the five heuristics. Parametric methods are defined as follows. Let x_{ik} be the number of copies of symbol i that have been already sequenced in the sequence of length k, k = 0, 1, … (asumme x_{i0} = 0); the symbol to be sequenced in position k + 1 is $i^* = \arg\max_i \left\{ d_i / \left(x_{ik} + \delta \right) \right\}$, where $\delta \in (0,1]$. Webster method and Jefferson method are parametric methods that use a δ value equal to 0.5 and 1, respectively. Insertion method is a recursive heuristic based on grouping symbols into fictitious symbols until only two fictitious symbols remains and then solving optimally the two-symbol cases.

In Herrmann (2007) was proposed an aggregation method based on grouping iteratively the symbols with the same number of copies to be sequenced into fictitious symbols and then applying a parametric method. This idea is extended in Herrmann (2008).

More complex algorithms based on metaheuristic schemes have also been proposed: one Cross-Entropy method (CE) algorithm (García-Villoria et al., 2007), two multi-start (MS) algorithms (García et al., 2006; Corominas et al., 2008), two greedy Randomized Adaptive Search Procedure (GRASP) algorithms (García et al., 2006; Corominas et al., 2008), eleven Particle Swarm Optimisation (PSO) algorithms (García et al., 2006; García-Villoria and Pastor, 2009a), one Electromagnetism-like Mechanism (EM) algorithm (García-Villoria and Pastor, 2009b), one Psychoclonal algorithm (García-Villoria and Pastor, 2008) and one Tabu Search (TS) algorithm (Corominas et al., 2009a).

All these metaheuristic algorithms, except the CE algorithm, have been tested on the same set of benchmark instances. The set is composed of 740 instances which were grouped into four classes (from CAT1 to CAT4 with 185 test instances in each class) according to their size. The instances were generated using the random values of D (total number of copies) and n (number of symbols) shown in Table 1. For all instances and for each symbol i = 1,…,n, a random value of d_i (number of copies to be sequenced of model i) is between 1 and $\left| (D-n+1)/2.5 \right|$ such that $\sum_{i=1..n} d_i = D$. All algorithms were coded in Java and all computational experiments were carried out on a 3.4 GHz Pentium IV with 1.5 GB of RAM.

Table 1. Uniform distribution for generating the D and n values

	CAT1	CAT2	CAT3	CAT4
D	U(25, 50)	U(50, 100)	U(100, 200)	U(200, 500)
n	U(3, 15)	U(3, 30)	U(3, 65)	U(3, 150)

Tables 1 and 2 show the average RTV values of the solutions obtained with the metaheuristics for 50 and 1,000 computing seconds, respectively (if there are more than one algorithm based on the same metaheuristic, only the results of the best of them are shown). The results are shown for the 740 instances and for each class of instances (CAT1 to CAT4).

The improvement of the results obtained with metaheurtistc for the benchmark instances during the last two years has been enormous. The ratio between the average of the RTV values obtained with the TS algorithm and the average of the RTV values obtained with the best PSO algorithm is around 0.05 (for 1,000 seconds of computing time). It could seem that the PSO proposed algorithms have a bad performance. However, they improved a lot the results obtained with the best heuristic proposed in Corominas et al. (2007), which is the Webster method. The Webster method obtains an RTV values average equal to 22,821.94; and considering the results by class, the average values obtained with the heuristic are 121.84, 933.11, 8,502.80 and 81,730.05 for CAT1, CAT2, CAT3 and CAT4 instances, respectively.

Anyway, these values should be taken only as reference information since the execution time needed by the heuristic is always less than one second per instance.

Table 2. Average RTV values for a computing time of 50 seconds

	Global	**CAT1**	**CAT2**	**CAT3**	**CAT4**
TS	202.42	10.30	22.40	109.38	667.59
Psycho	235.68	14.92	44.25	137.07	746.50
MS	2,106.01	11.56	38.02	154.82	8,219.65
GRASP	2,308.69	13.00	60.45	270.93	8,890.37
EM	3,747.05	19.14	54.54	260.79	14,653.72
PSO	4,625.54	16.42	51.34	610.34	17,824.04

Table 3. Average RTV values for a computing time of 1,000 seconds

	Global	**CAT1**	**CAT2**	**CAT3**	**CAT4**
TS	113.31	10.24	21.46	106.21	315.33
Psycho	161.60	14.90	39.90	122.38	469.23
MS	169.25	10.51	31.21	123.27	512.02
GRASP	301.90	11.56	50.45	227.50	918.10
EM	330.29	18.64	52.97	157.20	1,092.36
PSO	1,537.34	14.35	46.55	143.96	5,944.51

4. Conclusions and future research

Several heuristic and metaheuristics have been proposed in the literature for solving the RTVP but only two exact methods for the general case, both MILP models. At present, the practical limit for obtaining optimal solutions is around 40 copies to be sequenced. Note that MILP models use general software. Thus, an exact algorithm designed specifically for solving the RTVP may increase the size of instances that can be optimally solved. For example, a promising line of research is to develop a procedure based on the branch and bound algorithm for the RTVP.

Results of Table 1 show that simple metaheuristics based on applying local search coupled with a mechanism to escape from local optima (multi-start, GRASP and TS) obtain equal or better results than more complex metaheuristics (PSO, EM and psychoclonal).

A promising line of research based on the results of Table 1 is to design new algorithms based on other metaheuristics as, among others: Genetic Algorithm, Simulated Annealing (SA) and Variable Neighbourhood Search (VNS). Moreover, the existing TS algorithm (which is the best one) can be applied using different neighbourhood structures. At present, we are working on these lines of research.

The case of the multi-start algorithm proposed in Corominas et al. (2008) should be enhanced. This metaheuristic is based on, iteratively, obtaining a random solution and applying a local search procedure. The performance of this simple algorithm improves the performance of the PSO and the EM algorithms, which are much more complex. Another promising line of research, which nowadays we are also developing, is to use the general scheme of multi-start hybridized with VNS, TS or SA. That is, VNS, TS or SA can be applied instead of local search at each iteration of multi-start.

Finally, another non-exact line of research is to use hyper-heuristics. Hyper-heuristics are an emerging methodology in search and optimisation (Burke et al., 2003). Hyper-heuristic methods choose dynamically the most suitable (meta)heuristic among a set of them according to the state of the search of the solution.

References

Anily, S., Glass, C.A., Hassin, R. (1998). The scheduling of maintenance service. Discrete Applied Mathematics, Vol. 82, pp. 27-42.

Balinski, M.L., Young, H.P. (1982). Fair Representation: meeting the ideal of one man, one vote. Yale University Press, New Haven CT.

Balinski, M., Shahidi, N. (1998). A simple approach to the product rate variation problem via axiomatic. Operation Research Letters, Vol. 22, pp. 129-135.

Bautista, J., Companys, R., Corominas, A. (1996). A note on the relation between the Product Rate Variation (PRV) problem and the apportionment problem. Journal of the Operational Research Society, Vol. 47, pp. 1410-1414.

Bollapragada, S., Bussieck, M.R., Mallik, S. (2004). Scheduling Commercial Videotapes in Broadcast Television. Operations Research, Vol. 52, pp. 679-689.

Brusco, M.J. (2008). Scheduling advertising slots for television. Journal of the Operational Research Society, Vol. 59, pp. 1363-1372.

Burke, E.K., Kendall, G., Newall, J, Hart, E., Ross, P., Schulenburg, S. (2003). Hyper-heuristics: An Emerging Direction in Modern Search Technology. Chapter 16 in Handbook of Metaheuristics. Kluwer Academic Publishers.

Corominas, A., Kubiak, W., Moreno, N. (2007). Response time variability. Journal of Scheduling, Vol. 10, pp. 97-110.

Corominas, A., García-Villoria, A., Pastor, R. (2008). Solving the Response Time Variability Problem by means of Multi-start and GRASP metaheuristics. Special Issue of Frontiers in Artificial Intelligence and Applications on Artificial Intelligence Research and Development, Vol. 184, pp. 128-137.

Corominas, A., García-Villoria, A., Pastor, R. (2009a). Using Tabu Search for the Response Time Variability Problem. Technical report IOC-DT-P, Universitat Politècnica de Catalunya, Spain. Available at http://hdl.handle.net/2117/2804.

Corominas, A., Kubiak, W., Pastor, R. (2009b). Mathematical Programming Modeling of the Response Time Variability Problem. European Journal of Operational Research. doi: 10.1016/j.ejor.2009.01.014.

Dong, L., Melhem, R., Mosse, D. (1998). Time slot allocation for real-time messages with negotiable distance constrains requirements. Fourth IEEE Real-Time Technology and Applications Symposium (RTAS'98), Denver, CO. pp. 131-136.

García, A., Pastor, R., Corominas, A. (2006). Solving the Response Time Variability Problem by means of metaheuristics. Special Issue of Frontiers in Artificial Intelligence and Applications on Artificial Intelligence Research and Development, Vol. 146, pp.187-194.

García-Villoria, A., Pastor, R., Corominas, A. (2007). Solving the Response Time Variability Problem by means of the Cross-Entropy Method. Special Issue on Production Line Systems: Concepts, Methods and Applications of the International Journal of Manufacturing Technology and Management, to be published.

García-Villoria, A., Pastor, R. (2008). Solving the Response Time Variability Problem by means of a psychoclonal approach. Journal of Heuristics, doi:10.1007/s10732-008-9082-2.

García-Villoria, A., Pastor, R. (2009a). Introducing dynamic diversity into a discrete particle swarm optimization. Computers & Operations Research, Vol. 36, pp. 951-966.

García-Villoria, A., Pastor, R. (2009b). Solving the Response Time Variability Problem by means of the Electromagnetism-like Mechanism. International Journal of Production Research, doi: 10.1080/00207540902862545.

Han, C.C., Lin, K.J., Hou, C.J. (1996). Distance-constrained scheduling and its applications in real-time systems. IEEE Transactions on Computers, Vol. 45, pp. 814-826.

Herrmann, J.W. (2007). Generating Cyclic Fair Sequences using Aggregation and Stride Scheduling. Technical Report TR 2007-12, University of Maryland, USA. Available at http://hdl.handle.net/1903/7082.

Herrmann, J.W. (2008). Using Aggregation to Reduce Response Time Variability in Cyclic Fair Sequences. Technical Report TR 2008-20, University of Maryland, USA. Available at http://hdl.handle.net/1903/8382.

Kubiak, W. (1993). Minimizing variation of production rates in just-in-time systems: A survey. European Journal of Operational Research, Vol. 66, pp. 259-271.

Kubiak, W. (2009). Response Time Variability. Chapter 7 in Proportional optimization and fairness. Springer.

León, D., Corominas, A., Lusa, A. (2003). Resolución del problem PRV min-var. Technical report IOC-DT-I-2003-03, Universitat Politècnica de Catalunya, Spain.

Miltenburg, J. (1989). Level schedules for mixed-model assembly lines in just-in-time production systems. Management Science, Vol. 35, pp. 192-207.

Monden, Y. (1983). Toyota Production Systems. Industrial Engineering and Management Press, Norcross, GA.

Moreno, N. (2002). Solving the product rate variation problem (PRVP) of large dimensions as an assignment problem. Doctoral Thesis, DOE, ETSEIB, Universitat Politècnica de Catalunya, Spain.

Waldspurger, C.A., Weihl, W.E. (1994). Lottery Scheduling: Flexible Proportional-Share Resource Management. First USENIX Symposium on Operating System Design and Implementation.

Waldspurger, C.A., Weihl, W.E. (1995). Stride Scheduling: Deterministic Proportional-Share Resource Management. Technical Report MIT/LCS/TM-528, Massachusetts Institute of Technology, MIT Laboratory for Computer Science. Available at https://eprints.kfupm.edu.sa/67117.

Wei, W.D., Liu, C.L. (1983). On a periodic maintenance problem. Operations Research Letters, Vol. 2, pp. 90-93.