A new tool for analysing and reporting solutions for the RCPSP and MMRCPSP

José Coelho[1,3], Mario Vanhoucke[1,2], and Ricardo Amaro[3]

[1]Ghent University, Belgium
e-mail: jose.coelho, mario.vanhoucke@ugent.be

[2]Vlerick Business School, Belgium
[3]Universidade Aberta, Portugal
e-mail: jose.coelho@uab.pt

Keywords: RCPSP, datasets, reporting results.

1. Introduction

In the paper written by Vanhoucke and Coelho (2018), a new method is proposed to facilitate the reporting of results for the single- and multi-mode RCPSP. We have now extended this method with a website where researchers can download and upload solutions without much intervention, which is the topic of this abstract.

The new website does not want to replace the well-known existing libraries such as the PSPLIB proposed in Kolisch and Sprecher (1996), the MMLIB proposed in Peteghem and Vanhoucke (2014) or the generic OR-LIBRARY proposed by Beasley (1990), but rather serves as a complement. The website reports data about many benchmark datasets from the literature in a standardized way, and also provides the best LB/UB/optimal values, the best known solutions (start times of each activity), and also information about project indicators (network and resource indicators). We have saved exactly one result file for each run of a complete dataset, and the performance of the procedure used is calculated against the CPM lower bound as well against the current best LBs and UBs.

We also present two new datasets for the RCPSP. The first so-called NetRes set has already been proposed earlier in Vanhoucke and Coelho (2018), which is a large set of 30 activity instances that spans a wide range for the topological network structure. The second set is totally new and is proposed in Coelho and Vanhoucke (2020) and contains a small set of very hard instances with 20 to 30 activities. This so-called CV set contains the smallest possible instances that we could find for which no optimal solutions could be found using the fast and efficient branch-and-bound procedures from the literature.

In the remainder of this abstract, we will detail how the results are reported (Section 2). In Section 3, we describe how we will update the website tables with best known solutions for the RCPSP and the MMRCPSP. Section 4 provides an illustrative example of an experiment with NetRes. In Section 5, we show the diversity of the new CV dataset, and we conclude in Section 6.

2. Reporting new results

The method we propose for reporting new results is done using a single *data file* per dataset (in CSV format) rather than one file per instance, containing one line per instance. Each line contains all possible data for that instance, such that user can easily know the network and resource indicators for each instance in the set. The results are given in a singe *result file* (also in CSV format). Consequently, our method requires only a single file for each run and avoids the need to submit one result per instance. Not only the values of the LBs and UBs are made available, but also the obtained solutions by the author of the new algorithm (the start times of each activity), and these results can be interesting for other researchers.

A software tool – a client tool - was developed to allow users the read and modify the results file if they have found new and better results. In doing so, the LBs and UBs are checked automatically for errors or inconsistencies. If no errors are found, the results file is updated and a reference to the new paper for new solutions is given. The tool is also easy to use for selecting only a subset of instances of a dataset (e.g. only the open or closed files or the files with LBs x% from the best known UB) and the instances will be automatically be selected for the user in a so-called *instance file*.

The website is in solutionsupdate.ugent.be and is integrated in the projectmanagement.ugent.be/research/data. The website will maintain and update Table 2 and Table 3 of Vanhoucke and Coelho (2018) that contains data from several datasets. Other tables for other project scheduling problems can be also added in the future.

Even if no new results are found, the website can be used to submit results before the submission of the paper, and in doing so, the authors will have a confirmation that there results contains no inconsistencies (such as UBs lower than a strong LB). This can be done easily using the client tool, but when the results are put online, it also gives the reviewers the possibility to check.

3. Update of tables of BKS on RCPSP and MMRCPSP

In this section we report current results for the tables that we intend to keep updated in the website. *Table 1* displays the current best-known results for the RCPSP and is an update of Table 2 published in Vanhoucke and Coelho (2018). More specifically, we updated the table with the new CV set and the Patterson set. For the NetRes set, we also reported the results for the 1kNetRes set, which contains results for a subset of NetRes in which each instance is selected in steps of 1,000 (reducing the number of instances to e.g. 540,000 to 540 for the NR(SP) set). We can now compare the results with the table in the original paper to see the progress made in the last few years by many authors. The table reports the number of open instances in the PSPLIB (J60 to J120) have been reduced. This data is not easily detectable from the PSPLIB website as done in Table 1.

Table 1. Best-known results for the RCPSP

Dataset	Subset	#Instances	#Open	%CPM	GAP
CV		623	623	142.21%	3.3
RG30		1,800	116	39.27%	2.0
RG300		480	377	956.71%	35.2
DC1		1,800	0	26.57%	0.0
DC2		720	210	274.20%	7.6
PSPLIB	J30	480	0	13.38%	0.0
	J60	480	37	10.37%	6.3
	J90	480	66	9.43%	7.5
	J120	600	290	29.01%	8.0
NetRes	NR(SP) | 1k	540,000 | 540	25,591 | 12	78.8% | 72.9%	5.3 | 1.8
	NR(AD) | 1k	480,000 | 480	44,855 | 7	98.8% | 102.4%	5.6 | 1.1
	NR(LA) | 1k	720,000 | 720	246 | 0	58.4% | 58.9%	4.6 | 0.0
	NR(TF) | 1k	720,000 | 720	23,544 | 0	68.3% | 64.7%	6.4 | 0.0
	NR(RC) | 1k	540,000 | 540	10,333 | 0	66.3% | 71.6%	6.0 | 0.0
	NR(RU) | 1k	270,000 | 270	3,761 | 0	73.6% | 77.0%	9.3 | 0.0
	NR(VAR) | 1k	540,000 | 540	4,722 | 0	87.3% | 91.9%	4.3 | 0.0
Patterson		110	0	18.04%	0.0

As mentioned earlier, with this updated data, a reviewer can easily check whether some new results on the RCPSP are within a valid range by e.g. checking the percentage deviation of the LB over the CPM. Also, the sum of time units of both best lower bounds and best upper bounds is provided, and this indicator can be checked in the same way than the %CPM. This does not rule out the possibility of less credible researchers to invent and manipulate results, but prevents errors unwillingly made by the researchers. Nevertheless, the reviewer can also ask the researcher to submit a result file to the website, so the results can always be checked, even if there are no new LBs or UBs.

Table 2 displays the current best-known results for the MMRCPSP, and is an update of Table 3 published in Vanhoucke and Coelho (2018). The LBs are updated with the work of Stürck (2018), and compared with the version published in the paper, this new data lead to a larger number of instances closed in the MMLIB. This illustrates and highlights the importance of research on LBs as much as on UBs. Note that in the Boctor instances, the GAP between the UBs and the LBs is very high. This is mainly because no good LBs exist for these instances, since these instances do not contain non-renewable resource. The MMLIB site is no longer available, but the final UB values from 2018 are used in our website to guarantee we have used to most recent results. As for the RCPSP, a reviewer can also check new results.

Table 2. Best-known results for the MMRCPSP

Dataset	Subset	#Instances	#Open	%CPM	GAP
PSPLIB	J10 | J12 | J14 | J16 | J18 | J20	536 | 547 | 551 | 550 | 552 | 554	0 | 0 | 0 | 0 | 0 | 0	32% | 27% | 24% | 19% | 18% | 17%	0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0
	J30	552	245	12.28%	6.5
Boctor	Boct50	120	120	22.74%	52.6
	Boct100	120	120	22.91%	103.6
MMLIB	MMLIB50	540	95	22.29%	9.3
	MMLIB100	540	151	21.35%	10.8
	MMLIB+	3240	2439	78.77%	37.2

4. An example of an experiment with NetRes

The NetRes set was proposed in Vanhoucke and Coelho (2018), and the goal was to create a set with high diversity in terms of the project indicators, but also to provide a large number of instances available such that researchers can select subsets they need. Several analyses are done in the original paper, but we have select Table 5 of the original paper and replicate results in *Table 3* that measures the impact of the project indicators using the exact procedure of Demeulemeester and Herroelen (1992). An instance is considered hard if it could not be solved in 1 second, and the table shows the percentage of hard instances of each value of the project indicator (SP, AD, LA, TF, OS, RC and RS).

Table 3. Percentage of hard instances in NetRes depending on each project indicator

	SP	AD	LA	TF	OS	RC	RS
0-0,1	54%	-	4.9%	0.4%	59%	0%	7.4%
0,1-0,3	13%	1.5%	0.3%	0.5%	31%	9.1%	7.2%
0,3-0,5	0.2%	4.5%	0.1%	2.8%	1.3%	4.9%	1.3%
0,5-0,7	0%	13%	0.1%	8%	0.4%	3.8%	0.5%
0,7-0,9	0%	4.9%	0%	17%	0%	1.8%	0.1%
0,9-1	-	1.9%	-	24%	0%	1.4%	0%

As we can see in *Table 3*, most of the instances in this set are closed, but we can now visualize where the most complex instances are for each indicator. All the findings are more or less known (except for the new project indicators AD, LA and TF). For example, parallel networks (low SP and OS values) are harder to solve, and for the RC indicator, an easy/hard/easy phase transition is found, which confirms the results of Herroelen and De Reyck (1999). A similar effect is found for the AD indicator, and the indicators LA and RS provide more hard instances when the indicator is low. The TF indicator provides harder instances when it is high.

The *Table 3* is an example of an experiment that could not be easily done if no instances are available for all values of all these indicators. Vanhoucke et. al. (2016) have shown that most sets are not diverse enough, and only contain instances with values between 0 and 1 for some indicators, while others are largely ignored.

We expect that the NetRes set will be interesting for research where statistical tests are used extensively. A deeper study into the relation between a given project indicator and the performance of a solution procedure requires data that spans the full range of complexity. The client tool can help selecting the subset of instances necessary for such a study. Moreover, the large volume of instances with solutions could potentially be interesting for researcher using machine learning making use of the current best-known solutions on a large amount of data to train the data.

5. Diversity of dataset CV

Table 4 displays the distribution of the CV instance set for several project indicators used in Vanhoucke et. al. (2016). Recall that this set contains instances that are currently unsolvable. The table shows that this set of hard instances still contains instances with diversity in the network structure and resource constraints, and hence, not only contains instances with very parallel activities. All topological indicators are spread over a wide interval except for LA that is concentrated around values below 0,2. For the resource indicators, the RS is not very diverse and most of the instances have a value lower than 0,2. The diversity is higher for the other resource indicators, with RU greater than 2, RC between 0,2 and 0,5, RF greater than 0,8. This set is said to be very hard to solve, and researchers could focus their research time trying to solve these instances to optimality.

Table 4. Distribution of instances in CV dataset by several project indicators

#Activities	#Resources	CNC	OS	SP	AD
20-21 #4	1 #1	0-1 #405	0-0,1 #85	0-0,1 #232	0-0,2 #8
22-23 #18	2 #39	1-2 #176	0,1-0,2 #416	0,1-0,2 #316	0,2-0,4 #79
24-25 #41	3 #85	2-3 #24	0,2-0,3 #99	0,2-0,3 #62	0,4-0,6 #243
26-27 #95	4 #498	3-4 #6	0,3-0,4 #13	0,3-0,4 #12	0,6-0,8 #229
28-30 #465		4-8 #12	0,4-0,6 #10	0,4-0,5 #1	0,8-1 #64
LA	**TF**	**RC**	**RF**	**RU**	**RS**
0-0,2 #592	0-0,2 #28	0,2-0,3 #57	0,5-0,6 #9	1-2 #34	0-0,1 #533
0,2-0,4 #10	0,2-0,4 #48	0,3-0,4 #258	0,6-0,7 #13	2-3 #84	0,1-0,2 #87
0,4-0,6 #11	0,4-0,6 #138	0,4-0,5 #285	0,7-0,8 #153	3-4 #505	0,2-0,3 #3
0,6-0,8 #7	0,6-0,8 #220	0,5-0,6 #10	0,8-0,9 #245		
0,8-1 #3	0,8-1 #189	0,6-0,8 #13	0,9-1 #203		

The reason why we claim these instances are hard is that we have tried to solve these instances using 20 hours of CPU time for each instance with the procedure presented in Coelho and Vanhoucke, M. (2018), and we have reported the best found LB and UB. The percentage over the CPM of LBs is 129%, and this percentage increases to 142% when compared with the UBs, leaving enough space to find improvements for the 623 instances.

It is interesting to note that we have kept these instances as small as possible. Most instances contain 20 activities, and go up to 30 activities maximum, and some of them make use of only 1 renewable resource.

6. Conclusion

In this abstract, we present a new contribution to the academic community with a tool to keep the current results for the RCPSP and MMRCPSP updated at all times. The tool intends to save the latest results from all datasets in a standardized way, validates new results and provides performance indicators. We also provided a new large dataset NetRes that is diverse in several project indicators, allowing doing analyses for several project indicators, and a second new dataset CV with only small instances that are still not solved to optimality. We hope and believe that this tool and the new dataset can be used in new research studies, which can lead to entirely new solution procedures that can solve small but very hard instances to optimality.

References

Beasley J.E., 1990, "OR-Library: Distributing Test Problems by Electronic Mail ", Journal of the Operational Research Society, Vol. 41 (11), pp. 1069-1072.

Coelho, J., Vanhoucke, M. (2018). An exact composite lower bound strategy for the resource-constrained project scheduling problem. Computers & Operations Research, 93, 135–150.

Coelho, J., Vanhoucke, M. (2020). Going to the core of hard resource-constrained project scheduling instances. Computers & Operations Research, 121, 104976.

Demeulemeester E., W. Herroelen, 1992, "A branch-and-bound procedure for the multiple resource-constrained project scheduling problem", Management Science, Vol. 38, pp. 1803–1818.

Herroelen W., B. De Reyck, 1999, "Phase transitions in project scheduling", Journal of the Operational Research Society, Vol. 50, pp. 148–156.

Kolisch R., A. Sprecher, 1996, "PSPLIB – A project scheduling problem library", European Journal of Operational Research, Vol. 96, pp. 205-216

Peteghem V.V., M. Vanhoucke, 2014, "An experimental investigation of metaheuristics for the multi-mode resource-constrained project scheduling problem on new dataset instances", European Journal of Operational Research, Vol. 235, pp. 62-72

Vanhoucke M., J. Coelho, J. Batselier, 2016, "An overview of project data for integrated project management and control", Journal of Modern Project Management, Vol. 3 (2), pp. 6–21.

Vanhoucke M., J. Coelho, 2018, "A tool to test and validate algorithms for the resource-constrained project scheduling problem", Computers & Industrial Engineering, Vol. 118 (1), pp. 251 – 265

Stürck C., 2018, "Exakte Methoden und Matheuristiken für das Multi-Mode Resource-Constrained Project Scheduling Problem (Exact and matheuristic approaches for the multi-mode resource-constrained project scheduling problem)", Dissertation, Helmut-Schmidt-University, Hamburg