Project Scheduling Under Uncertainty

Tarun Bhaskar, GE Global Research, Bangalore, India; E-mail: tarun.bhaskar@ge.com
Manabendra Nath Pal, Indian Institute of Management Calcutta, Kolkata, India; E-mail: mnp@iimcal.ac.in

ABSTRACT
Scheduling is a very important part of the planning phase of project management. But the fact that it is a planning phase process makes it susceptible to uncertainty. This paper discusses uncertainty at different phases in project scheduling and then provides a method for handling uncertainty at the planning phase. We consider the project-scheduling problem with multiple resource constraints, discuss the uncertainty involved in the activity duration and its effect on the schedule. We propose a priority rule for a new schedule generation scheme, which takes care of the criticality of the activities and the randomness involved in the current and future activities. The method is demonstrated on a problem taken from the literature.

Keywords: Project Scheduling, Uncertainty, Robustness, Resource, Heuristic

1. INTRODUCTION
Scheduling is a decision-making process, which plays a crucial role in manufacturing as well as service industry. Scheduling mainly concerns the allocation of limited resources to jobs over time. This decision problem exists in lots of manufacturing and production systems as well as most of the information-processing units. According to the *Project Management Institute (PMI)* scheduling software is a "run-the-business workhorse" in many companies and it may improve success rates for them by up to 20% (Essex, 2005).

Several methods have been proposed for solving scheduling problems. These methods take the values of some parameters (such as resource and time required to complete each activity, available resources etc.) of the problem as the input and generate a schedule for the problem. In most of the cases, the parameters of the projects are estimated based on the domain knowledge and past experience of the manager. So there is some amount of uncertainty embedded in these estimates. The uncertainty can be at two levels:

1. At the planning level when the manager may not be very sure about the estimates of the parameter itself and
2. At the implementation level, when the manager gives a deterministic estimate of the parameter, but they change while executing the schedule.

In the first case the nature of the manager defines the type of input to the problem to a large extent. For example when the manager is risk averse, she would not be willing to give a deterministic estimates for the parameters. Therefore, she can either chose to give a vague estimate (for example a fuzzy number) or a random estimate or she may use some other measure, which represents uncertainty. In the second case the schedule, which was developed using the estimated values, becomes inefficient which leads to rescheduling the project several times. So a schedule, which is robust enough to take care of these changes to some extent, should be a preferred schedule. This paper is focused around the first aspect of uncertainty discussed above and discusses it for project scheduling problem.

The *Resource Constrained Project Scheduling Problem (RCPSP)* has been extensively studied in the literature. There can be different objectives of a project-scheduling problem, but minimization of project completion time is one of the most important objectives of a project-scheduling problem (Schonberger, 1981; Willis, 1985; Ulusoy and Ozdamar 1995). If we solve the problem considering this objective, time required by each activity becomes a very important input that is estimated based on past data and the experience of the project manager. So, activity duration estimates are susceptible to uncertainty.

The project-scheduling problem under resource constraints has been studied to a large extent but the literature is the field of scheduling with uncertainty is scarce (Demeulemeester and Herroelen, 2002). The research in this field can be classified into three main sections: criticality indices of activities and paths, probability distributions related to activity times and scheduling of activities. The criticality index (CI) of an activity is defined as the probability that the activity will be on the longest path (Dobin and Elmaghraby, 1985; Bowman 1995; Cho and Yum, 1997; Elmaghraby *et al.*, 1999). A detailed discussion on this topic has been provided in the review paper (Elmaghraby, 2000). The second topic of discussion in this area is the probability distribution of the activity and project completion time. The basic assumption of PERT network is that the activity time follows a beta distribution (Malcom *et al.*, 1959). Ginzburg (1988) suggested a new measure for the mean and variance of the distribution for activity duration

The third area is scheduling of activities under uncertainty (Malcom *et al.*, 1959; Schmidt and Grossmann, 2000; Pontrandolfo 2000). But very few studies considered resource constraint. Ginzburg and Gonik (1997) have proposed a simulation-based heuristic to solve the problem. At every decision point, they run the simulation to find out the criticality of each activity and then use a heuristic to allocate resources to activities. Their main concern was that the decision taken at a point is not only dependent on the past decisions taken, but also are dependent on the future decisions. So, at each decision point they run the simulation to find out the probability of a particular activity lying on the critical path. This repetitive process makes the method cumbersome and time consuming.

In this paper, we propose a heuristic method to solve the project-scheduling problem with multiple resource constraints, which has random activity duration. This method takes care of the concern of Ginzburg and Gonik (1997) without getting into a time consuming and cumbersome job of repetitive simulation. The heuristic proposed in this paper is a non-recursive method and gives an efficient solution to the problem. We use three different distributions (uniform, normal and beta) to model the uncertainty in the activity duration. We use a problem from the literature (Ginzburg and Gonik, 1997) to demonstrate our method.

The rest of the paper is structured as follows. In the next section we introduce the notations used in this paper and then we explain the problem in section 3. We discuss the proposed heuristic in section 4 and use that to solve a problem taken from the literature in section 5. Finally, we discuss the results, future research directions and then we conclude.

2. NOTATIONS
a_j Optimistic time required for activity j.
b_j Pessimistic time required for activity j.
σ_j The standard deviation of time for activity j.
j Activity number of the project. $j = 0, 1, 2, 3, ..., N+1$, where 0 and N+1 are dummy start and dummy end nodes respectively
t_j Duration of activity j, a random variable
M_j Maximum remaining path length of activity j, a random variable.
A_j Set of activities on the path of maximum length starting from j
A_0 Set of activities on the critical path
S_j Net standard deviation of the path associated to M_j. $S_j = \sqrt{\sum_{k \in A_j} \sigma_k^2}$ for all activities on that path.
r_{jk} Renewable resource of type k required to perform activity j. $k = 1, 2, 3, ..., K$.
R_k Total available resource of type k.
T Project completion time without resource constraint. This is the critical path length of the project based on the expected value of the time taken by each activity.
S Standard deviation of the critical path. $S = \sqrt{\sum_{k \in A_0} \sigma_k^2}$ for all activities on the critical path.

$\overline{R_k}$ Minimum resource of type k required to complete the project in time T assuming no other (K – 1) resource constraints.

T_k Time taken (based on the expected value of the activity time) to complete the project considering only the k^{th} resource constraint and no other (K – 1) resource constraints.

i Cycle number of a decision point. A decision point occurs either at the beginning of the project or when at least one of the running activities is completed.

R_{ik} Resource of type k available at cycle i.

C_i The set of activities which are ready to be scheduled in cycle i. This is the set of activities, which satisfies the precedence relationship.

C_{im} The subset m of the set C_i that can be formed taking as many activities as possible without violating any resource constraint. m = 1, 2, 3, ..., M. So, $C_i = \bigcup_{m=1}^{M} C_{im}$.

l_{im} Number of activities in the subset C_{im}.

Z_{im} The value of SPI (as explained in section 4) of subset m in cycle i

3. PROBLEM DESCRIPTION

The problem discussed in this paper is of scheduling the project under multiple resource constraints with ill-defined activity duration. Most of the time it is convenient to estimate the upper and lower bound of the activity duration. Based on these estimates, the problem is to find the expected completion time of the project. So the problem has the following properties:

- A set of activities
- Random activity duration, which is generated based on the lower and upper bounds
- Fixed resource requirement and availability. We consider only renewable resources.
- A fixed precedence relationship
- Preemption not allowed
- Objective of minimization of expected project completion time.

A mathematical formulation of the problem, as given in Ginzburg and Gonik (1997), formulates it as a stochastic optimization problem, which is a hard problem to solve. We propose a heuristic method to solve the problem.

4. THE PROPOSED HEURISTIC

Now we introduce the heuristic to solve the problem discussed above. This heuristic is based on a priority rule, which gives a priority list of set of activities at each decision point. A decision point occurs either at the beginning of the project, or when at least one of the running activities is completed, till the last activity is scheduled. At every decision point, a set of all activities (C_i) whose predecessors have been completed is formed. All possible subsets of this set are formed which satisfies the resource constraints. So, at every decision point, several subsets of activities compete for the same resource. We decide on the winning subset based on the priority rule. A deterministic version of the priority rule has been discussed (Bhaskar *et al.*, 2004), but the need of similar measure for probabilistic network is evident. We call this priority rule as Schedule Performance Index (SPI). The SPI is based on the following important points:

- The objective is to minimise the expected project completion time. So we need to take care of the most critical activities. So, we should schedule that subset of activities whose criticality factor is highest among all competing subsets.
- As discussed earlier, we agree with the concern of Ginzberg and Gonik (1997) that at every decision point we need to take care of the randomness of the duration of activities that has not yet been scheduled.
- The scarcity of the resources should be minimised. This can be done by scheduling, if possible, the subset of activities with maximum resource requirement at the earliest. By doing this we try to avoid any resource crunch in the future.

As discussed in the first two points above, we need to incorporate the criticality of the activities in SPI. The criticality factor in our work represents, for each subset of activities, its distance from the critical path taking care of the randomness in the activity duration. We use the concept of *Maximum Remaining Path Length* (MRPL) (Moder *et al.*, 1983) to take care of criticality. MRPL of a particular activity is defined as the length of the longest remaining path starting from that activity. This represents, at every decision point, how critical is a particular activity. If we add the net variance of the remaining path, it takes care of the randomness of the future activities. So, a factor that represents the proximity of subset of activities to the critical path, taking care of the randomness of future activities, can be given by:

$$\frac{M_j + nS_j}{T + nS}$$

Where n is a number representing the weightage given to the randomness of future activities. We discuss the effect of n on the final results in the later section. We divide the expression by ($T + nS$) to make the parameter less than 1, as ($T + nS$) is at least as much as the numerator and it remains constant throughout the project.

Now we consider the last point, i.e. regarding the utilization of the resources, discussed above. We schedule the subset that requires maximum amount of resources compared to other subsets. By doing this we minimise the probability of any resource crunch in future. This can be measured by the ratio of resource required to resource available. So we introduce the following factor in our priority rule:

$$\prod_{k=1}^{K} \left(\frac{\sum_{j \in c_{im}} r_k}{R_k} \right)$$

It is now a known fact that the complexity of the project scheduling under multiple resource constraints is not a linear function of the types of resources we use. So, we take the product over the types of resources in the problem.

The factors defined above are the measure of the criticality and the resource management of the subsets respectively. These measures are calculated at all decision points. So, this gives the local perspective of the problem at the decision point. We now introduce some global measures of the problem which remains constant throughout the problem and which represents the overall perspective of the problem. We have calculated the weight of each resource type in the problem, which represent the criticality of that type of resource. This can be given by the probability of a resource crunch of that type of resource during the whole time span of the project. This can be measured by the ratio of the resource available and the resource required to complete the project in the minimum possible time. This ratio gives us the criticality of that project. So we use this ratio as the power to represent the weight of a particular type of resource. This can be represented as:

$$p_k = \frac{\overline{R_k}}{R_k}$$

Similarly, to measure the overall probability of time overrun as the weight of the time factor explained above can be given as follows:

$$q = \frac{\max_k (T_k)}{T}$$

The value of T and T_k in the above expression is based on the expected value of the time taken by each activity.

So, the final expression for SPI, which is used in finding the winning subset, can be obtained by combining all these factors and it is represented as:

$$Z_{im} = \prod_{k=1}^{K} \left(\frac{\sum_{j \in C_{im}} r_{jk}}{R_{ik}} \right)^{p_k} \cdot \frac{1}{l_{im}} \sum_{j \in C_{im}} \left(\frac{M_j + nS_j}{T + nS} \right)^{q}$$

$\overline{R_k}$ and T_k can be calculated using Burgess and Killewbrew Algorithm (Burgess and Killebrew, 1962) and Brook's Algorithm (Bedworth, 1973) respectively. It is evident from the expression of SPI that the whole expression becomes zero if the resource of one or more type required by the activities in a particular subset C_{im} at any decision point is zero. To take care of these situations, we introduce some remedies for this. For any subset of activities, C_{im}, at a decision point i, if one or more resource type (but not all) are not required, then we can postpone the activities as discussed above. So in this case where the value of

$$\left(\frac{\sum_{j \in c_m} r_{jk}}{R_k} \right)$$

goes to zero, we replace the value by a very small positive real number ∂ (say 0.001). Where as in the case where there is no requirement of any type of resource

Now we use the above priority rule (SPI) to develop the heuristic algorithm for scheduling the project with uncertain activity times. To find the final schedule, we follow the following steps:

- Find the critical path length and the standard deviation of the critical path based on the mean of the distribution considered for random activity time.
- Use the Burgess and Killewbrew Algorithm to calculate $\overline{R_k}$
- Use the Brook's Algorithm to find T_k.
- Calculate the values of p_k and q.
- At each decision point:
  - Generate the random variate based on the parameters of the assumed distribution. The distributions considered in this paper and its parameters have been discussed in section 5.
  - Find the activities whose predecessor activities have been completed and populate the set C_i
  - Make the subsets C_{im} from the elements of the set C_i, which satisfies the resource constraints.
  - Calculate Z_{im} for each of these subsets
  - Schedule the subset with maximum value of Z_{im}.

This algorithm takes care of the uncertainties of the activities, which have not been scheduled at a particular point in time. Using this algorithm we schedule the activities till all the activities are completed and find out the completion time.

5. EXPERIMENT AND RESULTS

In this section we test our heuristic algorithm using an example taken from Ginzburg and Gonik (1997). The project under consideration has 36 normal and two dummy (start and end nodes) activities. These activities require renewable resource of only one type. The number of resources available is 50. The details of time required to complete an activity, the precedence relationship and the resource requirements are given in Table 1. The time required to complete an activity is not well known, therefore the optimistic and pessimistic time estimates are provided in the data.

For this experiment, we use three different standard probability distributions to generate the project activity duration. The justification of a particular distribution for PERT type of network is outside the scope of this paper. The distributions used in this study are:

1. A beta distribution in the interval $[a_j, b_j]$;
2. A uniform distribution with the range $[a_j, b_j]$
3. A normal distribution with mean as $(a_j + b_j)/2$ and variance as $[(b_j - a_j)/6]^2$

We generate random variates based on the above distributions and use them as the activity durations. To take care of the randomness and to get the average characteristic of the solution, we run the algorithm for 1000 times for each distribution and take the average of those 1000 runs as the project completion time. The results for different n are given in Table 2. The average project completion time for each run is rounded off to the just higher integer in case of non-integer completion time. But the average completion time, which is stated below, is the exact average of those integer completion times.

Table 1. Initial data of the test project

Activity no. (j)	r_j	a_j	b_j	Successors
0	0	0	0	1, 2, 3, 4, 5
1	16	40	60	6, 7
2	15	35	70	10, 11
3	18	25	35	12
4	19	30	45	13
5	10	26	33	14
6	18	9	15	8, 9
7	24	38	50	27, 28
8	25	10	18	26
9	16	16	24	27, 28
10	19	30	38	17, 18
11	20	18	22	26
12	18	25	32	24, 25
13	15	31	45	17, 18
14	16	58	78	15, 16
15	17	35	45	20, 21, 22
16	19	25	35	23
17	21	35	60	19
18	24	30	50	20, 21, 22
19	13	35	42	35
20	16	20	30	33
21	12	14	21	34
22	14	15	20	35
23	16	30	42	33
24	15	28	40	30, 31
25	13	22	28	32
26	14	20	35	29
27	18	16	24	29
28	22	15	22	30, 31
29	10	13	18	36
30	18	27	38	36
31	16	35	55	37
32	17	20	30	34
33	19	25	27	37
34	20	17	38	36
35	15	38	55	37
36	24	12	22	37
37	0	0	0	-

The results stated above gives the expected completion time of the project and they reveal lots of interesting facts. The expected completion time in case of beta distribution is less than that of other distributions. So, the assumption of a beta distribution gives an optimistic estimate of the completion time compared to other distributions. As we increase the value of n, the completion time increases in most of the cases because the value of n represents the weightage we give to the randomness of the future activities. So, higher the weightage given to the randomness of the future activities, higher is the time of completion.

The result gives an indication that for higher value of n, the chance of completion of the project within the estimated time should be higher. We try to analyze this observation by doing one more experiment. For each distribution we calculate the project completion time by taking the $\mu_j + n\sigma_j$ as the deterministic activity duration of each activity and used the proposed method for scheduling. We calculate the number of instances of project (in case of random activity times) where project completion time lies within the time estimated by the deterministic

Table 2. Project completion time for different distributions

n	Uniform	Beta	Normal
0	422.22	410.25	421.53
1	436.86	412.62	438.21
2	433.69	424.82	437.56
3	453.11	418.56	448.58
4	468.73	432.56	462.86

case with $\mu_j + n\sigma_j$. This would give us an indication of chance of completion of project within the time estimated by taking $\mu_j + n\sigma_j$ as the activity duration. The preliminary results show that around 50% of the instances lie within the calculated time estimate in case of $n = 0$ for almost all the distributions (in case of beta distribution this value was 43%). As we increased the value of n, the chance of completion increases. For $n = 4$, the chance of completion of project was 97.1%, 97.5% and 95% for uniform, normal and beta distributions respectively. These are only indicative results, which show that there may be a relationship between the value of n and the probability of completion of project within some time. A theoretical study in this regard needs to be done and we consider it as an interesting area for future research.

6. CONCLUSIONS

In this paper we have discussed uncertainty involved in project scheduling under resource constraints. We have discussed uncertainty involved in the planning as well as the implementation phase of the problem and have proposed a method to take care of uncertainty in the planning phase. We have proposed a new efficient heuristic for project scheduling under multiple resource constraints and random activity duration. The heuristic is non-recursive and does not require simulation at each decision point. It also takes care of the concern of the researchers that the decision taken at any decision point should also be a function of the randomness associated with the future activities. The method is tested on a problem taken from the literature. The results also show that beta distribution, compared to the other two distributions, gives an optimistic measure of project completion time. In this work we have considered randomness only in activity duration.

This work can be extended in two directions. The result shows that the project completion time increases with increase in the value of n. It would be interesting to find out the exact value of the probability of completion of project as a function of n. Finding the most appropriate probability distribution or even a bound on the probability would be of interesting. The second extension can be considering uncertainty in the resource requirement and the resource availability as well. This study becomes more interesting as it is clear that the uncertainties in resource and activity duration are correlated.

REFERENCES

Bedworth D.D., Industrial Systems: Planning, Analysis and Control. The Ronald Press Co., New York, 1973.

Bhaskar T., Pal R. and Pal M.N., Resource Time Ratio Exponent Technique (RETIREXT):An Efficient Non-recursive Heuristic for Project Scheduling under Multiple Resource Constraints. Ninth International Workshop on Project Management and Scheduling (PMS 2004): 2004, pp. 92-95.

Bowman R.A., Efficient Estimation of Arc Criticalities in Stochastic Activity Networks. Management Science. 41(1), 1995, pp. 58-67.

Burgess R. and Killebrew J.B., Variation in Activity Level on a Cylindrical Arrow Diagram. Journal of Industrial Engineering. 13(2), 1962, pp. 76-83.

Cho J.G. and Yum B.J., An Uncertainty Importance Measure of Activity in PERT Networks. International Journal of Production Research. 35, 1997, pp.2737-2757.

Clark C.E., The PERT Model for the Distribution of an Activity. Operations Research. 10, 1962, pp.405-406.

Demeulemeester E.L. and Herroelen W.S., Project Scheduling: A Research Handbook (International Series in Operations Research and Management Science) 2002, First edition. Springer.

Dodin B.M. and Elmaghraby S.E., Approximating the criticality indices of the Activities in PERT Networks. Management Science. 31(2), 1985, pp.207-223.

Elmaghraby S.E., On Criticality and Sensitivity in Activity Networks. European Journal of Operational Research. 127, 2000, pp.220-238.

Elmaghraby S.E., Fathi Y. and Taner M.R., On Sensitivity of Project Variability to Activity Mean Duration. International Journal of Production Economics. 62, 1999, pp.219-232.

Essex D.E. Master the Clock. PM Networks: The Professional Magazine of the Project Management Institute. June, 2005.

Golenko-Ginzburg D., On the Distribution of Activity Time in PERT. Journal of Operations Research Society. 39(8), 1988, pp.767-771.

Golenko-Ginzburg D. and Gonik A., Stochastic Network Project Scheduling with Non-consumable Limited Resources. International Journal of Production Economics. 48(1), 1997, pp.29-37.

Moder J.J., Phillips C.R. and Davis E.W., Project Scheduling with PERT, CPM and Precedence Diagramming. 1983, Van Nostrand Reinhold Company.

Pontrandolfo P., (2000). Project Duration in Stochastic Networks by the PERT-path Technique. International Journal of Project Management. 18, 2000, pp.215-222.

Schonberger R.I., Why Projects are Always Late: A Rational Based on Simulation of a PERT/CPM Method. Interfaces. 11(5), 1981, pp.66-70.

Ulusoy G. and Ozdamar R., A Heuristic Scheduling Algorithm for Improving the Duration and Net Present Value of a Project. International Journal of Operations and Production Management. 15(1), 1995, pp.89-98.

Willis R.J., Critical Path Analysis and Resource Constrained Project Scheduling: Theory and Practice, European Journal of Operational Research, 21, 1985, pp.149-155.

Related Content

Architecture as a Tool to Solve Business Planning Problems
James McKee (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 573-586).*
www.irma-international.org/chapter/architecture-as-a-tool-to-solve-business-planning-problems/183772

Mechanical Transmission Model and Numerical Simulation Based on Machine Learning
Pan Zhang (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-15).*
www.irma-international.org/article/mechanical-transmission-model-and-numerical-simulation-based-on-machine-learning/318457

Modified LexRank for Tweet Summarization
Avinash Samueland Dilip Kumar Sharma (2016). *International Journal of Rough Sets and Data Analysis (pp. 79-90).*
www.irma-international.org/article/modified-lexrank-for-tweet-summarization/163105

An Empirical Analysis of Antecedents to the Assimilation of Sensor Information Systems in Data Centers
Adel Alaraifi, Alemayehu Mollaand Hepu Deng (2013). *International Journal of Information Technologies and Systems Approach (pp. 57-77).*
www.irma-international.org/article/empirical-analysis-antecedents-assimilation-sensor/75787

Dynamics in Strategic Alliances: A Theory on Interorganizational Learning and Knowledge Development
Peter Otto (2012). *International Journal of Information Technologies and Systems Approach (pp. 74-86).*
www.irma-international.org/article/dynamics-strategic-alliances/62029