MULTIPLE RESOURCE LEVELING IN CONSTRUCTION SYSTEMS
THROUGH VARI-ETC(U)

R. C. LEACHMAN

UNCLASSIFIED
MULTIPLE RESOURCE LEVELING IN CONSTRUCTION SYSTEMS THROUGH VARIATION OF ACTIVITY INTENSITIES

by

ROBERT C. LEACHMAN
MULTIPLE RESOURCE LEVELING IN CONSTRUCTION SYSTEMS
THROUGH VARIATION OF ACTIVITY INTENSITIES

by

Robert C. Leachman
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

AUGUST 1980

Submitted for publication in the Naval Research Logistics Quarterly. This paper won first prize in the 1980 George E. Nicholson Student Paper Competition of the Operations Research Society of America.

This research was supported by the Office of Naval Research under Contract N00014-76-C-0134 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.
MULTIPLE RESOURCE LEVELING IN CONSTRUCTION SYSTEMS
THROUGH VARIATION OF ACTIVITY INTENSITIES

Robert C. Leachman

Operations Research Center
University of California
Berkeley, California 94720

Office of Naval Research
Department of the Navy
Arlington, Virginia 22217

Approved for public release; distribution unlimited.

Resource Leveling
Critical Path Analysis
Construction Systems
Manpower Allocation
Nonlinear Programming

(SEE ABSTRACT)
ABSTRACT

The resource leveling problem for a construction system producing a stream of output units is considered. The system is modeled using a critical-path-analysis activity network, from which an extended network is developed for an integrated planning effort of all output units. Activity intensity variables are defined which measure activity demand rates for resources and consequent activity durations for the production of each output unit. A heuristic approach consisting of an iterative non-linear programming procedure is presented which computes activity durations (intensities) for the minimization of resource capacity costs subject to meeting construction due dates. The application to a major ship overhaul is described, in which the procedure was used to level workloads of the various labor-trade shops.
Multiple Resource Leveling in Construction Systems Through Variation of Activity Intensities

Robert C. Leachman
University of California, Berkeley

1. Introduction

In construction systems such as shipbuilding or aircraft assembly, one finds the construction of a stream of large output units whose resource requirements interface through time. Typically, there are various progress and completion due dates on each output unit, and the problem arises of minimizing peak demands for various resources subject to meeting these due dates.

In the usual formulation of the resource leveling problem for a construction project, a critical path analysis of the project is made, in which component activities have fixed durations and fixed rates of resource utilization. The problem considered is to schedule the activities so as to minimize resource loading peaks. The difficult, combinatoric nature of the problem has resulted in an emphasis on heuristic techniques [2] [5] [6] [10].

However, in many construction systems, activities are "rubber band" in nature; that is, there is a feasible range of intensity of resource assignments to each activity, resulting in a range of possible activity durations. In this paper, the leveling of the various resource histories through variation in the intensity of activity resource loading is investigated.

2. The Model

2.1. Extended Network For Output Streams

We consider a system with component activities \( i = 1, \ldots, N \) utilizing \( K \) exogenous resources to produce \( M \) output units. We suppose the construction of an individual output unit may be represented by a CPM node scheme activity network, in which the nodes represent the activities, and the arrows indicate work flow dependencies.* The network is to be sufficiently

* In most presentations of CPM/PERT, the roles of nodes and arcs are reversed from the roles used here. Both formats are used in practice. See [6].
general so as to model any of the output units, where work requirements may vary from output unit to output unit.

Sherrard and Mehlick [7] have used such an approach to perform PERT simulations of a series of construction projects. In their paper, the projects are performed according to a basic activity network, but activity durations vary from project to project due to varying work requirements and learning effects. They demonstrate the effects on completion times of individual projects.

However, to dynamically model resource loading, an integrated planning effort is required, capturing the interaction of resource demands among the various output units. We consider an extended network whose nodes are denoted by \( i_m \), where \( i \) denotes the activity and \( m \) denotes the output unit. The immediate predecessors of node \( i_m \) in the extended network are defined as follows:

(i) \( i_{m-1} \), except when \( m=1 \);

(ii) \( j_m \), for all activities \( j \) which are immediate predecessors of \( i \) in the activity network.

Using the extended network, we envision the construction of \( M \) output units as a single project.

2.2. Activity Intensities

The dynamic activity analysis of production, in which activity output rates and resource loads are measured in terms of intensity functions and technical coefficients, is due to Shephard et al. [8] [9]. These authors developed the analysis for continuous flow production systems governed by linear inequalities. We adapt it here for the extended network format.

The rate of application of resources to each activity \( i \) working on each output unit \( m \) is modeled linearly in terms of a strictly positive intensity variable \( z_{im} \) taken with technical coefficients defined as follows:

(a) \( c_{im} \) = fraction of output unit \( m \) completed by activity \( i \) per unit intensity of \( i \), working on \( m \), per unit time, \( i=1,\ldots,N, m=1,\ldots,M \).

(b) \( a_{imk} \) = amount of exogenous resource \( k \) applied to activity \( i \) per unit intensity of \( i \), working on output unit \( m \), per unit time, \( i=1,\ldots,N, m=1,\ldots,M, k=1,\ldots,K \).

These technical coefficients vary from output unit to output unit, reflecting varying work requirements as well as learning curve behavior.
The intensity assignment \( z_m \) to node \( i_m \) of the extended network is maintained at a constant value between start and finish times for \( i_m \), but the particular level of intensity is a decision variable. In this way, the time assignment to each node \( i_m \) becomes a variable, given by

\[
i_m = \frac{1}{z_m c_m}, \quad i=1, \ldots, N, \quad m=1, \ldots, M.
\]

For each node \( i_m \) of the extended network, there is a maximum intensity assignment \( \tilde{z}_m \), termed an intensity bound. These bounds reflect available workspace and other conditions peculiar to each activity.

An initial selection of node intensities allows the ordinary CPM scheduling computations (see [6]) of early (late) start and finish times and total activity slack, using the node durations (1). In particular, operating all activities at their intensity bounds, ie,

\[
z_m = \tilde{z}_m, \quad i=1, \ldots, N, \quad m=1, \ldots, M,
\]

yields a minimum project time span. We shall refer to such a project policy as the greedy plan.

The intensity \( z_m \) is said to be critical if the time assignments given by (1) are such that node \( i_m \) is on a critical path in the extended network. If, on the other hand, node \( i_m \) has slack, it is an indication that activity \( i \) working on output \( m \) could be operated at a lower intensity, using less resources per unit time, without slipping the project schedule. In general, the total slack of a given slack path through the extended network can be allocated among the activities on that path, whereby the reduced intensity assignments would become critical. In fact, all nodes can be made critical by eliminating all slack from the slack subnetwork, ie, the subnetwork of slack paths of the extended network.

In following sections, we attack the resource leveling problem using a technique for allocating the slack of the slack subnetwork to reduce node intensity assignments.

3. Resource Levelling

3.1. Formulation

We suppose the various due dates for the project are expressed as late finish times for a subset of the nodes of the extended network. It is assumed that all terminating nodes, ie, nodes with no followers, have such dates. For simplicity of exposition, it is also assumed that all initial
nodes, i.e., nodes with no predecessors, have early start time 0. Denoting the set of initial nodes by \( I \) and the set of terminating nodes by \( D \), there is thus a maximum time length allowable for each path

\[ P_{\alpha_m, \beta_m} \]

through the extended network starting with a node \( \alpha_m, \in I \) and ending with a node \( \beta_m, \in D \). This length is given by the late finish time of \( \beta_m \), which we denote by \( LF(\beta_m) \).

For the resource leveling problem, we consider the non-storable resources required by the construction system; these are the resources which cannot be accumulated, such as the services per unit time of various machines and labor. It is assumed that such resources cannot be "hired and fired" with the variation of project resource loading histories, and consequently resources adequate to meet peak demands must be maintained for the life of the project.

Let \( NS \leq K \) be the number of non-storable resources, and renumber (if necessary) such resources \( k=1, \ldots, NS \). Assuming linear capacity costs, let \( p_k \) be the cost of maintaining a unit of capacity of resource \( k, k=1, \ldots, NS \), for the life of the project, where such capacity is required to meet peak demands.

To simplify the calculation of project resource loads, we consider a bar graph approximation. We take a discrete time grid \( t = 0, 1, 2, \ldots, T \), where \( T \) is the last project due date, and the grid is fine enough so that

\[ \exists m \in \mathbb{Z}, m \geq 1, i=1, \ldots, N, m=1, \ldots, M. \]  

That is, each node requires at least one time unit duration. For a given set of node durations

\[ \left\{ t_m \right\}, i=1, \ldots, N, m=1, \ldots, M, \]

one calculates the project loading of resource \( k \) during \([\tau, \tau+1)\) for an early start schedule to be

\[ \sum_{i=1}^{N} \sum_{m=1}^{M} \delta(i, m, \tau) \frac{a_{im}}{c_m} \frac{1}{t_m}. \]

where

\[ \delta(i, m, \tau) = \begin{cases} 
1 & \text{if } \tau - t_m < ES(i_m) \leq \tau, \\
0 & \text{if not},
\end{cases} \]
Here, $ES(i_m)$ denotes the early start of node $i_m$. The delta functions defined by (5) are seen to indicate whether node $i_m$ operates during $[\tau, \tau+1)$ or not.

We consider the problem of minimizing costs for non-storable resource capacities, subject to meeting project due dates. Let $X_k$, $k=1, \ldots, NS$, be variables representing the peak demands for non-storable resources encountered by the project. The problem may then be formulated as follows.

Minimize $\sum_{k=1}^{NS} p_k X_k$

subject to

$$R(1) \sum_{tm \text{ on } P_{\alpha_m \beta_m}} t_{im} \leq LF(\beta_m),$$

for all $\beta_m \in D$, and all $\alpha_m \in I$, and for all paths $P_{\alpha_m \beta_m}$ through the extended network starting at $\alpha_m$, and ending at $\beta_m$.

$$R(2) \sum_{i=1}^{N} \sum_{m=1}^{M} \delta(i,m,\tau) \frac{a_{im}}{c_{im}} \frac{1}{t_{im}} - X_k \leq 0, \ k=1, \ldots, NS, \ \tau=0,1, \ldots, T-1,$$

where $\delta(i,m,\tau)$ is defined by (5).

$$R(3) \ t_{im} \geq \frac{1}{z_{im}c_{im}}, \ i=1, \ldots, N, \ m=1, \ldots, M.$$

$$R(4) \ t_{im}, \ X_k \geq 0, \ i=1, \ldots, N, \ m=1, \ldots, M, \ k=1, \ldots, NS.$$

Constraints $R(2)$, for which one must relate node start times to the time grid $\tau=0,1, \ldots, T-1$, are a source of much difficulty. The literature contains efforts to handle constraints of this kind using integer variables, making the problem too large to solve for all but very short time horizons $T$. A convenient survey of such efforts is Herroelen [3]. We shall resort to a heuristic approach, as described in the next section.
3.2. Peak Pricing Procedure

The procedure presented here arises from the observation that, in many projects, activities in the very early and very late stages of the project, while requiring time to perform, are not resource intensive. Such activities might include tests, inspections, system checkouts, etc. Thus resource demand peaks for such projects are somewhat inflexible as to time of occurrence, restricted to mid-project ranges of the time grid.

A sequence of problems are considered in which only nodes operating in time regions of peak resource demand in the solution at hand are charged for resource capacity. Following optimization of each problem, inspection is made to see if peak regions have shifted, and, if so, to redefine the set of nodes to be charged for capacity in another iteration. The procedure is presented in algorithmic form as follows.

Step 0. Compute the greedy plan, i.e., set node intensities and time assignments as in (2) and (1), respectively. Compute a CPM forward pass schedule for these time assignments, and if all project due dates are not met, stop; the project is infeasible. Otherwise, go to Step 2.

Step 1. Compute a CPM forward pass schedule for the node time assignments of the current solution.

Step 2. Compute the project resource demand histories for the current solution using (4) and (5). From these histories, for each resource \( k = 1, \ldots, NS \), identify time regions of peak demand to be "priced," whereby activities operating in such regions will be charged for capacity in this iteration. The region specification is somewhat arbitrary, but should include time intervals experiencing peak and near-peak demand, and exclude intervals of low resource demand. Time intervals excluded from peak regions are likely to experience higher levels of resource demand in the next solution. See Figure 1 for an example of region pricing.

Based on these peak region definitions, compute

\[
\delta(i, m, \tau) = \begin{cases} 
1 & \text{if } \delta(i, m, \tau) = 1 \text{ for some } \tau \text{ such that } \tau, \tau + 1 \text{ belongs to the peak region of } k. \\
0 & \text{if not.}
\end{cases}
\]

\( i = 1, \ldots, N, \ m = 1, \ldots, M, \ k = 1, \ldots, NS. \) (6)

If the pricing scheme (6) has not changed since last iteration, stop; otherwise, solve the following program.
Activity Durations:

Pricing Assignments:

\[ \delta(A, 1,k) = \delta(C, 1,k) = \delta(B, 3,k) = 0 \]

\[ \delta(A, 2,k) = \delta(B, 2,k) = \delta(C, 2,k) = 1 \]

FIGURE 1
EXAMPLE OF PEAK REGION PRICING
Minimize \[ \sum_{i=1}^{k} \sum_{m=1}^{N} \sum_{m=1}^{M} \delta(i, m, k) p_k \frac{a_{mk}}{c_{im}} \frac{1}{(c_{im} c_{im})^{-1} + s_{im}} \]

subject to

\[ M(1) \sum_{i=1}^{k} \sum_{m=1}^{N} \sum_{m=1}^{M} s_{im} \leq LF(\beta_{m_1}) - \sum_{i=1}^{k} \sum_{m=1}^{N} (c_{im} c_{im})^{-1} \]

for all \( \beta_{m_1} \in D \), and all \( \alpha_{m_1} \in I \), and for all paths \( P_{\alpha_{m_1}} \beta_{m_2} \) through the extended network starting at \( \alpha_{m_1} \) and ending at \( \beta_{m_2} \).

\[ M(2) s_{im} \geq 0, \quad i=1, \ldots, N, \quad m=1, \ldots, M. \]

Step 3. Set \( t_{im} = s_{im} + (c_{im} c_{im})^{-1} \), \( i=1, \ldots, N, \quad m=1, \ldots, M \), and go to Step 1.

The procedure is seen to terminate when peak regions stabilize. We shall refer to the final set of node time assignments generated by the procedure as the resource leveled plan.

In each iteration a nonlinear objective is minimized subject to linear constraints. The program is formulated in terms of slack allocation variables

\[ \left\{ s_{im} \right\}, \quad i=1, \ldots, N, \quad m=1, \ldots, M. \]

so that it may be conveniently restricted to those nodes belonging to the slack subnetwork for the greedy plan calculated in Step 0. Constraints \( M(1) \) would then effectively consider only paths in the slack subnetwork.

There is no particular justification for starting the procedure with an early start schedule of the greedy plan; a late start schedule could also be used. At the end of each iteration of the procedure, however, all nodes are made critical. This can be seen by noting that any slack on a path from an initial node to a terminating node could be allocated to those activities operating in a region of peak resource demand, thereby decreasing the objective. Thus the procedure looks only among project plans in which all intensities are critical.

It is to be emphasized that the procedure is a heuristic, and it is not difficult to construct a simple example with very flexible peak region locations which will cause the procedure to cycle two solutions. However, for such a case, the pricing scheme (6) may be readily adapted to allow the immediate escape from such a cycle by simply pricing peak regions from both solutions.
3.3. Applications and Computational Experience

The Peak Pricing Procedure was applied to a network consisting of 151 activities modeling a ship overhaul. Since only a single overhaul was being modeled, the activity network served as the extended network. Exogenous inputs to the activities consisted of labor hours of twelve different trade shops. Unit capacity costs were developed from average wage and overhead rates for each shop.

Computation of the greedy plan established a minimum project time span of 370 working days. However, such a plan was characterized by serious demand peaks for various trade shops. The problem of reducing the costs of manpower capacity without slipping the project span was addressed.

The slack subnetwork under consideration was reduced to that consisting of activities whose start times and slack possibly allowed them to operate in the middle time range of the project where resource peaks could occur. The reduced network for application of the Peak Pricing Procedure then consisted of 33 activities contained in 48 slack paths.

The computer code used in the Peak Pricing Procedure applied the Frank-Wolfe linearization algorithm, which computed a solution within a 1% tolerance of the optimum for each subproblem. (See [1] for a description of the algorithm.) Peak resource regions were defined to be all work days where resource demand was within 4% of peak demand.

After four iterations, peak regions stabilized, so that no further improvement was possible for the Procedure. Each iteration had required about 10-15 seconds effective time on a CDC 6400 computer. Figure 2 compares work loads of the mechanical shop for the greedy production plan and the resource-leveled production plan. As can be seen, considerable improvement in shop loading has been made. Overall, a 38% improvement in the objective function was made between the initial greedy plan and the last iteration of the pricing procedure.

The Peak Pricing Procedure was also applied to a two-output unit extended network of 42 nodes, developed from a 21 activity subnetwork of the aforementioned ship overhaul network. Figure 3 compares workloads of the shipboard mechanical shop for the late start schedules of the greedy plan and the resource leveled plan. In this application, a similar number of iterations, computer run times, and objective function improvement were experienced as in the previous application. It appears that considerably larger problems could be handled by the procedure.
MECHANICAL SHOP WORK LOAD BEFORE AND AFTER RESOURCE LEVELING
FIGURE 3

SHIPBOARD MECHANICAL SHOP WORK LOAD - TWO SHIPS
3.4. Intensity Lower Bounds

Considerable replanning of activities to lower intensity levels was made by the pricing procedure to obtain the load leveling discussed in the applications. In many systems, there may be practical lower limits on activity intensity, i.e., intensity lower bounds

\[ z_m, i=1, \ldots, N, m=1, \ldots, M. \]

An important special case is an activity which can only be performed "one way," i.e., upper and lower intensity bounds are equal.

Such limitations can be incorporated into the pricing procedure in the form of upper bounds on node time assignments. The constraints \( M(1) - M(2) \) need only be supplemented by additional constraints as follows.

\[ M(3) \quad s_m \leq (z_m c_m)^{-1} - (z_m c_m)^{-1}, \quad i=1, \ldots, N, \quad m=1, \ldots, M. \]

In this situation, all node durations might not be critical in the solutions generated by the Peak Pricing Procedure, as the node duration upper bounds may be such that the absorption of all slack in the network is impossible. If such is the case, different schedules of the resource leveled plan may produce different resource loading histories, and so a two-phased procedure is suggested. First, application of the Peak Pricing Procedure is made to determine the resource leveled plan. Second, application of an heuristic scheduling algorithm such as developed in references [2], [5] or [10] could be made to search for an improved rescheduling of the early start schedule generated by the pricing procedure.

3.5. Activity Load Leveling Program

In the special case that no two activities share the same resource, an alternative non-linear programming formulation allows optimal solutions to be obtained for the resource leveling problem, which in this case amounts to an activity load leveling problem. Let \( NS(i), i=1, \ldots, N \), denote the number of non-storable resources required by activity \( i \). Then the program is to

\[
\text{Minimize} \quad \sum_{i=1}^{N} \sum_{k=1}^{NS(i)} p_k \frac{1}{B_{ik}}
\]

subject to

\[ S(1) \quad \sum_{m=m}^{n} s_m \leq LF(\beta_m) - \sum_{m=m}^{n} (z_m c_m)^{-1}. \]
for all $\beta_{m_2} \in D$, and all $\alpha_{m_1} \in I$, and for all paths $P_{\alpha_{m_1}\beta_{m_2}}$ through the extended network starting at $\alpha_{m_1}$ and ending at $\beta_{m_2}$:

$$S(2) \quad \frac{a_{m_1}}{c_{m_1}} B_i - s_m \leq (\bar{c}_{m_1} c_{m_1})^{-1}, \ i=1, \ldots, N, \ m=1, \ldots, M, \ k=1, \ldots, NS(i).$$

$$S(3) \quad s_m, B_i \geq 0, \ i=1, \ldots, N, \ m=1, \ldots, M, \ k=1, \ldots, NS(i).$$

Optimal node time assignments are then obtained by setting

$$t_{m_1} = s_m + (\bar{c}_{m_1} c_{m_1})^{-1}, \ i=1, \ldots, N, \ m=1, \ldots, M.$$  \hspace{1cm} (7)

Here, each variable $B_i$ is the inverse of the peak demand by activity $i$ for its resource $k, k=1, \ldots, NS(i), i=1, \ldots, N$. Constraints $S(2)$ define these variables. The problem has been formulated with these variables so as to obtain a constraint set of linear inequalities, making for ease of solution. Other variables and constants are defined as in the Peak Pricing Procedure. Lower intensity bound constraints $M(3)$ may also be introduced into this problem.

4. Conclusion

It has been shown how critical path analysis may be generalized into a production planning and scheduling tool for construction systems. The Peak Pricing Procedure (or Activity Load Leveling Program) applied to an extended network can be used to plan resource capacities and schedule activity operations subject to given production due dates.

The extended network model of a system producing a stream of output units constitutes a discrete transfer model of production. That is, completed units of work by an activity which are required as inputs by other activities are transferred unit by unit on an event basis. The considerable detail provided in modeling work in process in this fashion may prove to be intractable unless these units are "large," as is the case for the shipbuilding example which motivated this model. In models of other kinds of production, a considerable number of output units are producible per unit time, and transfers of work in process may reasonably be viewed as continuous flows governed by linear inequalities, as in references [8] and [9]. A forthcoming paper by the author will discuss production planning with continuous flow models.
Acknowledgements

This research was supported by the Office of Naval Research under Contract N00014-76-C-0134 with the University of California. This paper was adapted from [4], the author’s doctoral dissertation. The cooperation of the Puget Sound Naval Shipyard is gratefully acknowledged.

References