Cost-Based Directed Scheduling : Part I, An Intra-Job Cost Propagation Algorithm*

JAE KYEONG KIM
School of Business Adminstration,
KyungHee University
(jaek@khu.ac.kr)

MIN SOO SUH
Sales and Production Planning,
Hyundai Steel Company
(mssuh@hyundai-steel.com.kr)

Constraint directed scheduling techniques, representing problem constraints explicitly and constructing schedules by constrained heuristic search, have been successfully applied to real world scheduling problems that require satisfying a wide variety of constraints. However, there has been little basic research on the representation and optimization of the objective value of a schedule in the constraint directed scheduling literature. In particular, the cost objective is very crucial for enterprise decision making to analyze the effects of alternative business plans not only from operational shop floor scheduling but also through strategic resource planning. This paper aims to explicitly represent and optimize the total cost of a schedule including the tardiness and inventory costs while satisfying non-relaxable constraints such as resource capacity and temporal constraints. Within the cost based scheduling framework, a cost propagation algorithm is presented to update cost information throughout temporal constraints within the same job.

Key Words : Constraint Directed Scheduling, Cost Based Scheduling, Cost Propagation Algorithm

Received : October 2007 Accepted : December 2007 Corresponding author : JAE KYEONG KIM

1. Introduction

Constraint directed scheduling techniques, representing problem constraints explicitly and constructing schedules by constrained heuristic search, have been successfully applied to real world scheduling problems that require to satisfy a wide variety of constraints(Fox 1983 ; Smith

* This work was supported in part by Professor Overseas Research Support Enterprise of LG Yonam Cultural Foundation.

& Cheng, 1993 ; Nuijten et al., 1994 ; Le Pape, 1994 ; Caseau & Laburthe, 1994 ; Sadeh & Fox, 1996 ; Beck et al., 1997a, 1997b ; Easton & Goodale, 2005). Most of the work has focused on single performance measures that are non-decreasing in job completion times, which include mean flow time, mean lateness, percentage of jobs tardy, and mean tardiness. In such performance measures, the focal point has been put on the most critical constraint in terms of feasibility that has the higher risk of breaking resource constraints. However, in more complicated scheduling environments in industries such as Just-In-Time production, a cost-based scheduling concept has been required to explicitly represent and optimize the total cost of a schedule including the tardiness and inventory costs while satisfying non-relaxable constraints such as resource capacity and temporal constraints. In the constraint-directed scheduling literature, there has been little basic research on the cost-based scheduling with a few exceptions(Fox, 1983 ; Sadeh, 1991 ; Sun et al., 2003). Therefore, it has been not much known how to propagate cost information in each search state throughout a whole network of temporal constraints, and to estimate the cost impact of a scheduling decision on the total schedule cost. Without such *cost propagation algorithm*(*CPA*), decision making in scheduling heuristics relies upon local, incomplete cost information, resulting in poor schedule performance from the overall cost minimizing objective. In this paper, we present *intra-job* CPA. Whenever there is a change in cost information of an activity in a job in the process of scheduling, the intra-job CPA updates cost curves of other activities connected through temporal constraints within the same job.

The organization of this paper is as follows. Section 2 describes the target problem where cost components are introduced and represented. CPAs to create and update cost curves of each activity, and to propagate cost information along temporal constraints are explained in detail in Section 3 and Section 4. Section 5 describes concluding remarks and future research directions.

2. The Problem

The basic problem constraints studied in this paper come from the job shop scheduling problem where each job has n totally ordered activities in respective m resources. Each activity requires exclusive use of a singe resource for some processing duration. There are two types of constraints in this problem :

1) Precedence constraints between two ordered activities stating that if activity A is preceding activity B in the total order then A must be executed before B (that is, A → B).

2) Disjunctive resource constraints specifying that no two activities requiring the same resource may execute at the same time.

Depending upon the characteristics of the orders received and/or shop floor, we can consider a variety of cost components such as the tardiness, inventory, machine usage, labor, and changeover costs. When a large penalty is imposed on violating the due date of a job, the tardiness cost influences mostly on the schedule cost of the job. On the other hand, if the opportunity cost to manufacture a job is higher, the inventory and machine usage costs play a key role in determining the job schedule cost. The objective of the cost optimization problem is to reduce as much as possible the total cost of schedule while satisfying non-relaxable constraints such as precedence and resource constraints.

Notation

A_i : an activity i

$J(A_i)$: the job of activity A_i

$dd_{J(A_i)}$: the due date of job J_i

du_i : the processing duration of activity A_i

est_i : the earliest possible start time of activity A_i

lsi_j : the latest possible start time of activity A_i

eft_i : the earliest possible finish time of activity A_i

$$eft_i = est_i + du_i$$

lft_i : the latest possible finish time of activity A_i

$$lft_i = lst_i + du_i$$

STD_i : a feasible time window consisting of possible start times of activity A_i,

$$STD_i = \{est_i, est_i + 1, \cdots, lst_i\}$$

$A\text{-}p_i$: an activity that directly precedes A_i

$A\text{-}f_i$: an activity that directly follows A_i

$S_{fol}(A_i)$: a set of activities following activity Ai within job $J(A_i)$

$S_{pre}(A_i)$: a set of activities preceding activity Ai within job $J(A_i)$

S_J : a set of activities belonging to job J

We will omit the subscript unless there is the possibility of ambiguity. To represent cost components for each job, we adopt Sadeh's cost model having two cost components as follows :

1) Tardiness cost : When job J is scheduled to finish at time C_J past its due date dd_J, it is charged with a unit tardiness cost utc_J. The total tardiness cost of job J is :

$$TC_J = utc_j \times Max(0,\ C_J - dd_J)$$

2) Inventory cost : A unit inventory holding cost $uic(A_i)$, is charged to each activity A_i from its start time $st(A_i)$, until the last activity of its job is finished and delivered to customers. The total inventory cost of job J, is :

$$IC_j = \sum_{A_i} uic(A_i) \times [Max(C_j,\ dd_j) - st(A_i)]$$

The tardiness costs of jobs includes tardiness penalties and lost profit caused by not satisfying customers' preferences on due dates, while the inventory holding cost of a job represents an opportunistic cost for the expenses consumed in processing the job on the shop floor by workforce, machine usage, and material handling. The more a job stays at factory, the higher its inventory cost becomes. Therefore, the minimum cost of a job can be obtained when the last activity of the job is scheduled to finish on the due date of the job, and all preceding activities within the job are scheduled compactly without any time gap between activities. Starting from the due date of a job, we can calculate the initial optimal start time of each activity belonging to the job. As the scheduling process proceeds, a job may not meet its due date or may create a time gap between activities within the job due to the contention for resource capacity. This affects the optimal cost of a job and optimal start times of each activity accordingly. The next Section explains in detail how to update optimal start times of each activity when there is a change in the feasible time window or precedence constraint of an activity, and to propagate cost information through temporal constraints within the same job.

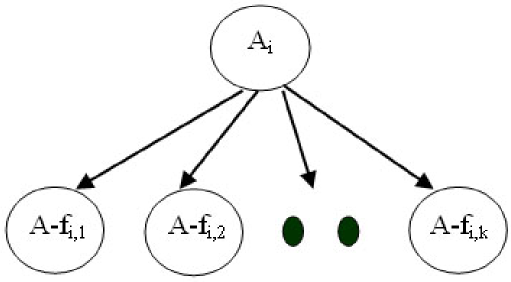
3. Cost Propagation Algorithms

As a starting point, an initial optimal start time of each activity is calculated under the assumption that there is no resource constraint, *i.e.*, all the possible start times of an activity are available. Let ost_A and oft_A denote the optimal start and finish times of activity A within job J, respectively. For the last activity A_L of job J, the initial optimal start time $ost(A_L)$ is :

$$ost(A_L) = Max(dd_J, \sum_{A_k \in S_J} du_{A_k}),$$
$$ost(A_L) \in [est(A_L), lst(A_L)]$$

Suppose activity A_i has a set of direct downstream activities, $\{A\text{-}f_{i,1}, A\text{-}f_{i,2}, \cdots, A\text{-}f_{i,k}\}$. The initial optimal start time $ost(A_i)$ is :

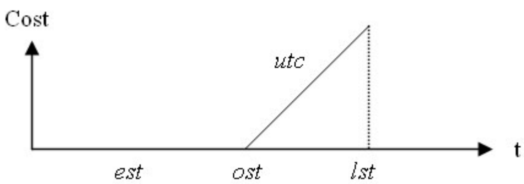
$$ost(A_i) = \underset{k}{MIN}(ost(A-f_{i,k}) - du(A_i)),$$
$$ost(A_i) \in [est(A_i), lst(A_i)]$$



If each activity within a job can be assigned at its optimal start time without any constraint violation, then the job will be finished exactly on its due date in the shortest staying time, making the total schedule cost of the job to be minimum. When there is a contention between activities for the same resource within a specific time interval, heuristic decision-making may be required to resolve the resource contention, and to adjust the feasible time window of contending activities. Whenever there is a change in the earliest and latest start times of an activity by implied or heuristic decision-making, there is a possibility for an activity to change its optimal start time and optimal cost. To maintain and propagate cost information for each activity in each search state of scheduling, we need two CPAs: intra-job CPA and inter-job CPA. The intra-job CPA is applied to activities belonging to the same job, while the inter-job CPA further visits other jobs connected through temporal constraints. This research presents the intra-job CPA, but the intra-job CPA will be presented in the subsequent research.

4. Intra-Job Cost Propagation Algorithm

Within a job, the intra-job CPA creates and maintains two types of cost curves(CCs) for each activity : tardiness CC and inventory CC. The initial tardiness CC of activity A starts from the optimal start time $ost(A)$, and adopts the unit tardiness cost $utc_J(A)$ as its initial positive slope as shown in [Figure 1].



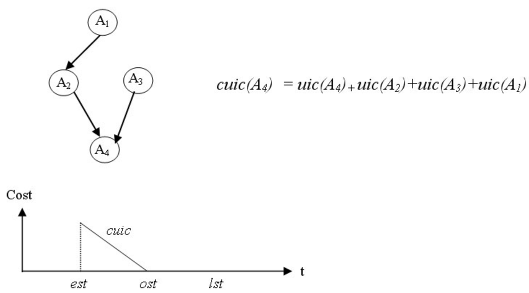
[Figure 1] An initial tardiness CC of an activity

In the beginning, each activity within a job has the same slope of tardiness CC. If an activity shrinks its feasible time window to the extent that a new earliest start time becomes greater than the optimal start time, then the job of that activity cannot meet its due date, and becomes tardy, resulting in the tardiness cost of the job.

The initial inventory CC also starts from the optimal start time of an activity but has, as an initial slope, a negative slope called *cumulative unit inventory cost* (*cuic*). The $cuic_A$ of activity A is obtained by summing up the unit inventory costs of all activities upstream of activity A, i.e.,

$$cuic_A = uic_A + \sum\{uic(A_k)|A_k \in S_{pre}(A)\}$$

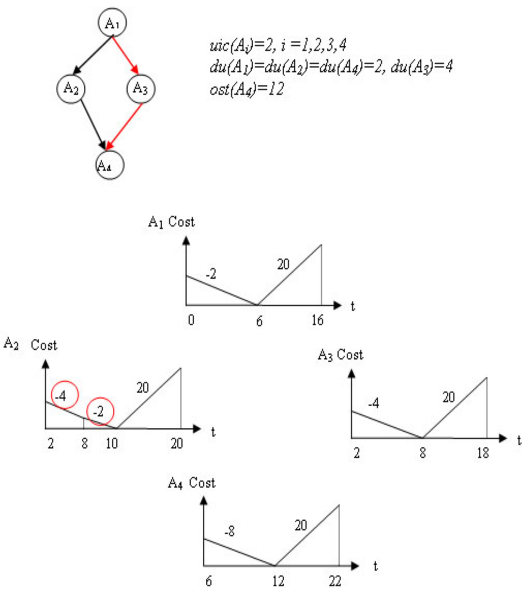
[Figure 2] shows an inventory CC of activity A_4 having three upstream activities. Suppose that activity A_4 has a new latest start time less than its current optimal start time. By the precedence constraint in the activities upstream of A_4, all preceding activities should also reduce their latest start times, resulting in additional inventory cost respectively. Therefore, the initial slope of inventory CC for an activity accounts for the total inventory costs including its upstream activities.



[Figure 2] An initial invertory CC of activity A_4

If there is a time gap in the optimal start time between upstream and downstream activities, the inventory CC should be split into two pieces due to the slope change as shown in [Figure 3] where for the time interval[8,10] in activity A_2, the unit inventory cost of activity A_1 is subtracted from the initial inventory slope of activity A_2, *cu-*

$ic(A_2)$, increasing the inventory CC slope from -4 to -2. This split of an inventory CC does not happen to the activities residing on the critical path which is the longest path to the last activity of a job. In other words, if an activity not on the critical path has an optimal start time gap with its upstream activity, its inventory CC should be split. In [Figure 3], only activity A_2 is on the non-critical path.



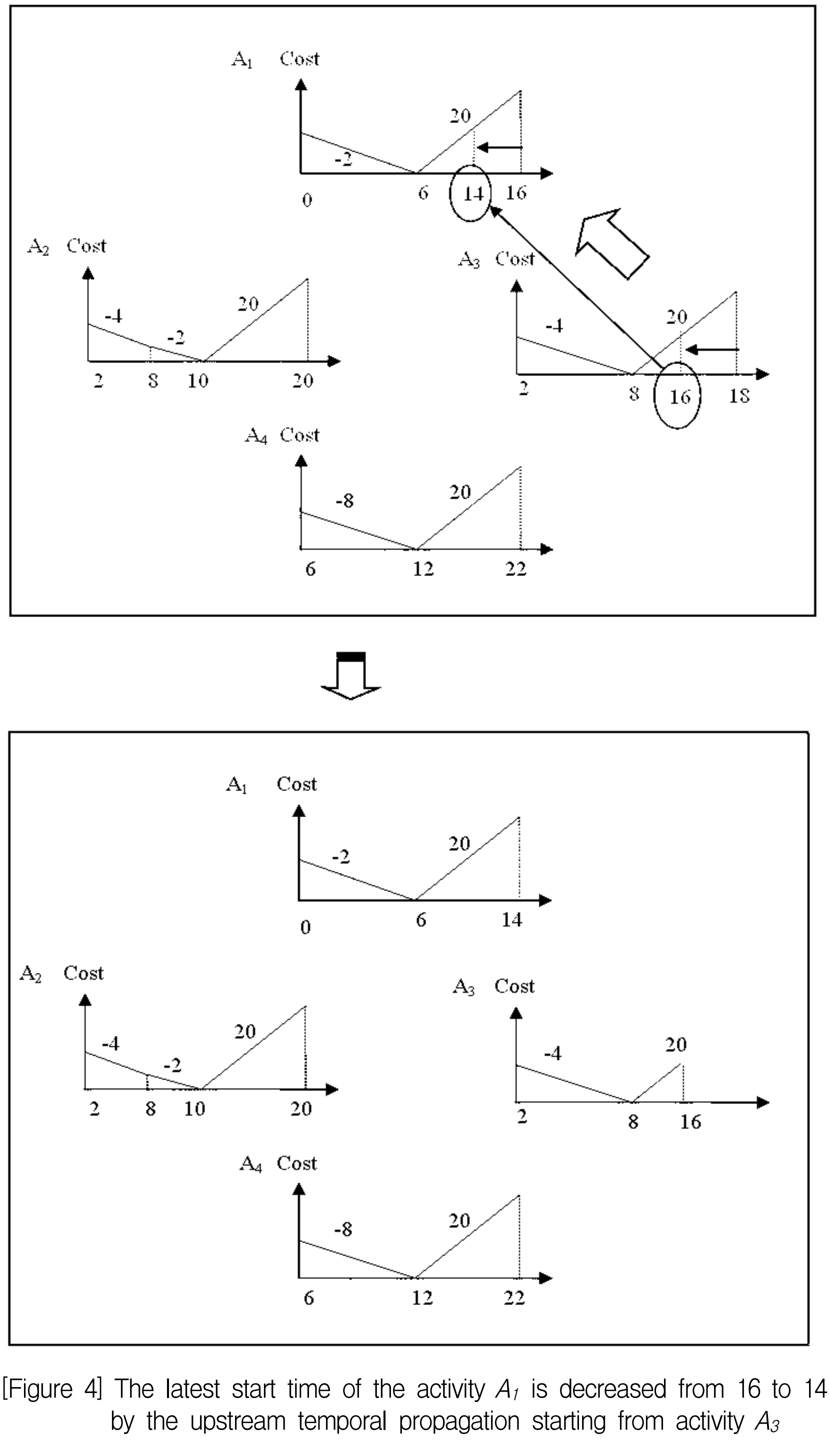
[Figure 3] Split of the inventory CC of activity A_2 due to a time gap between $ost(A_1)$ and $ost(A_2)$

Two types of events trigger cost propagation across the paths of temporal constraints : reduction of the latest start time and increase of the earliest start time.

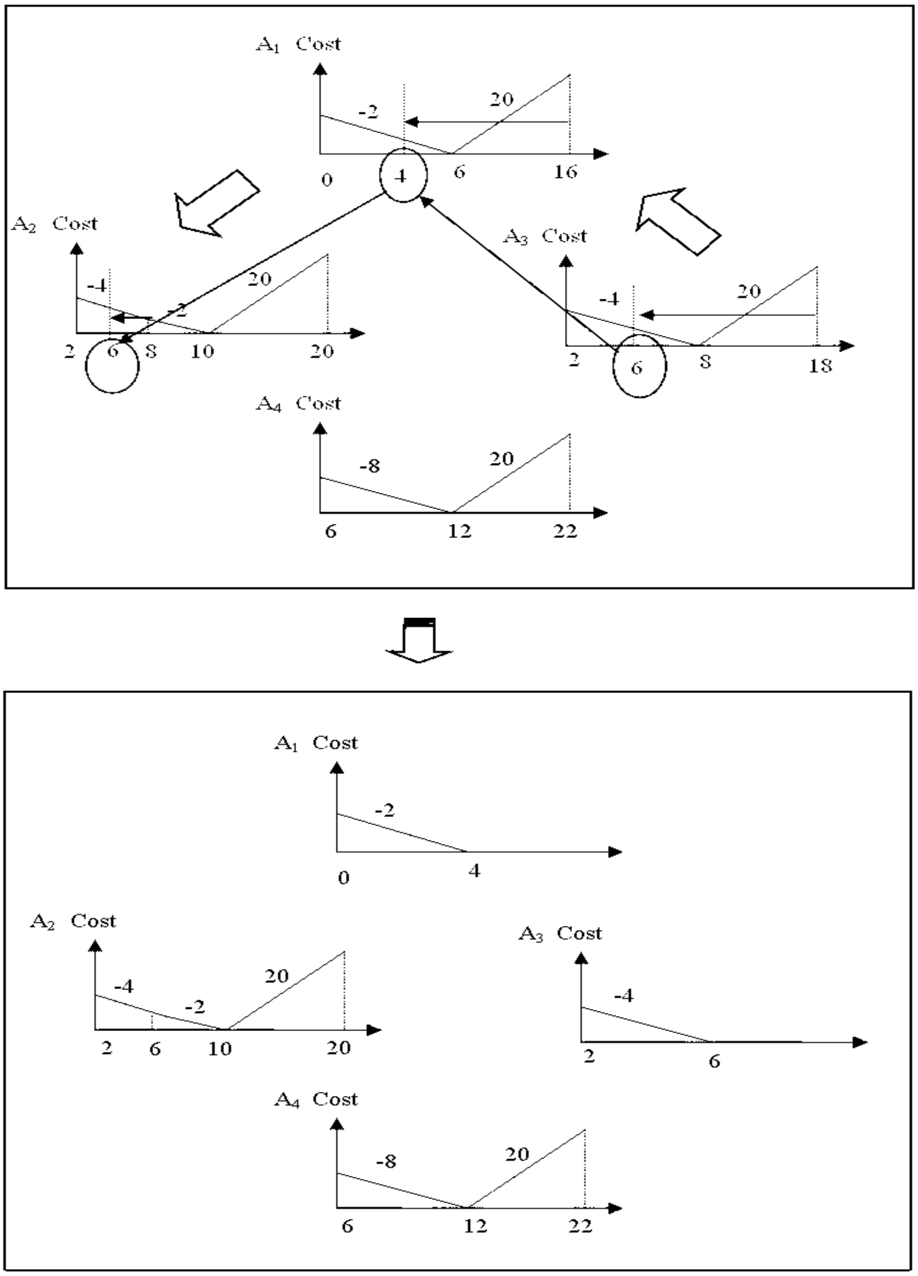
4.1 Reduction of the latest start time

When the latest start time of activity A is reduced from lst'_A to lst'_A, the domain of CCs in activity A shrinks accordingly, and cost information propagates up and down to activities upstream and/or downstream of activity A.

Upstream Propagation : By temporal propagation, the latest start time of each activity $A-p$ direct upstream of activity A may be reduced to $lst'_{A\text{-p}}$ where $lst'_{A\text{-p}} = Min(lst'_A - du_{A\text{-}p}, lst_{A\text{-}p})$. If the optimal start time of activity $A\text{-}p$ does not change, i.e., $lst'_{A\text{-}p} \geq ost_{A\text{-}p}$, only the domain of the tardiness CC for $A\text{-}p$ shrinks without no slope change. [Figure 4] shows that when activity A_3 changes its latest start time from 18 to 16, its upstream activity A_1 reduces the latest start time to 14 as a result of temporal propagation. If the optimal start time of activity $A\text{-}p$ changes due to temporal propagation, there might be a risk of a time gap between the optimal start time of activity $A\text{-}p$ and the corresponding time points in its downstream activities on the non-critical path, requiring downstream cost propagation as shown in [Figure 5] where, before cost propagation, activity A_2 had the time point, 8, corresponding to the optimal start time of its upstream activity A_1. That time point in activity A_2 should move left to 6 since the optimal start time of A_1 moved from 6 to 4 by the upstream propagation starting from activity A_3, resulting in that the slope of inventory CC for activity A_2 in the time interval[6,8] increases from −4 to −2.

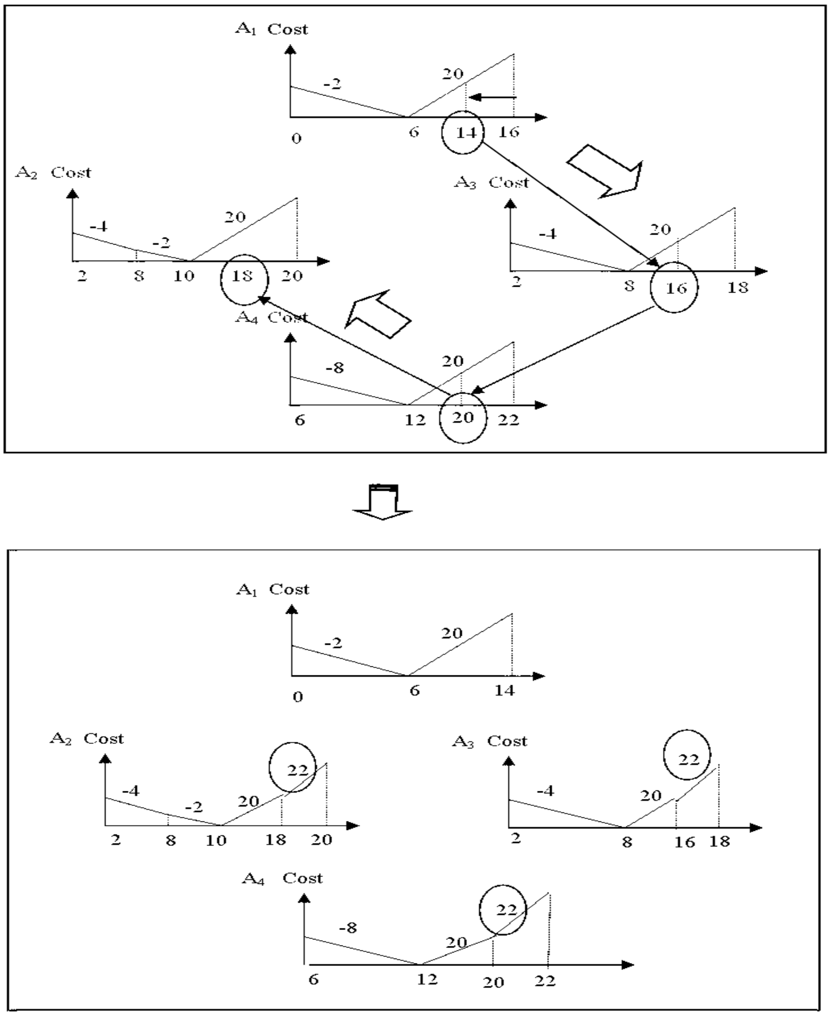


[Figure 4] The latest start time of the activity A_1 is decreased from 16 to 14 by the upstream temporal propagation starting from activity A_3



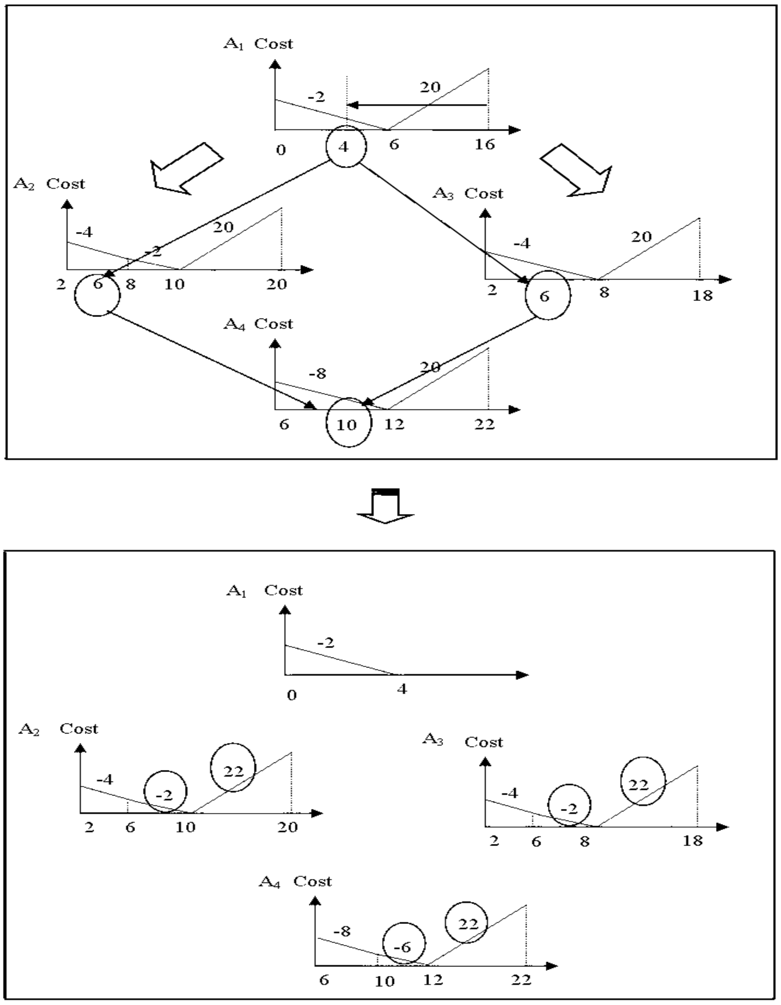
[Figure 5] The slope of inventory CC in activity A_2 increases from −4 to −2 for the time interval[6, 8] due to the decrease of the optimal start time of activity A_1

Downstream Propagation : There is no change in the feasible time window of activities downstream of activity A, but the slopes of CCs in the downstream activities may be increased. When there is no change in the optimal start time of activity A, cost propagation goes down only to the downstream activities on the critical path as shown in [Figure 6] where the downstream cost propagation starts from activity A_1. Suppose that activity A_2 moves to the right beyond the time point 16 after activity A_1 decreased its latest start time to 14. Then, activity A_1 cannot be scheduled compactly with activity A_2 after the time point 16 in activity A_2, resulting in additional inventory cost of activ-



[Figure 6] When the lastest start time of activity A_1 decreases from 16 to 14, the slope of tardiness CC for activity A_3 in the time interval[16, 18] increases from 20 to 22 by adding the unit inventory cost of activity A_1.

ity A_1. Therefore the slope of tardiness CC for activity A_2 in the time interval[16,18] increases from 20 to 22 by adding the unit inventory cost of activity A_1. This slope increase of tardiness CCs is repeated as the propagation goes down to downstream activities. In [Figure 6], activity A_4 has another upstream activity A_2, and in this case, the cost propagation goes up to activity A_2 to update the slope of tardiness CC for activity A_2. When there is a change in the optimal start time of activity A, cost propagation goes down to not only the activities on the critical path, but also to the activities on the non-critical path as shown in [Figure 7].

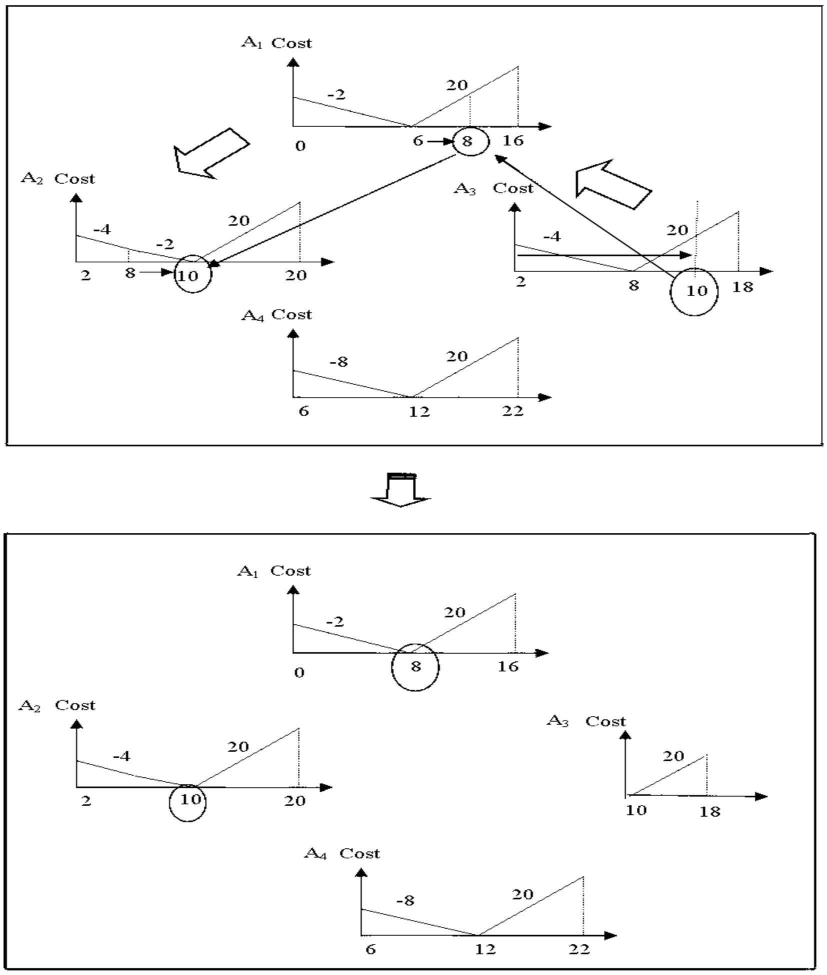


[Figure 7] When the optimal start time of activity A_1 decreases from 16 to 14, cost propagation goes down to not only activity A_3 on the critical path, but also to activity A_2 on the non-critical path.

4.2 Increase of the earliest start time

When the earliest start time of activity A is increased from est_A to est'_A, the domain of CCs in activity A shrinks accordingly, and cost information propagates along temporal constraints of activity A.

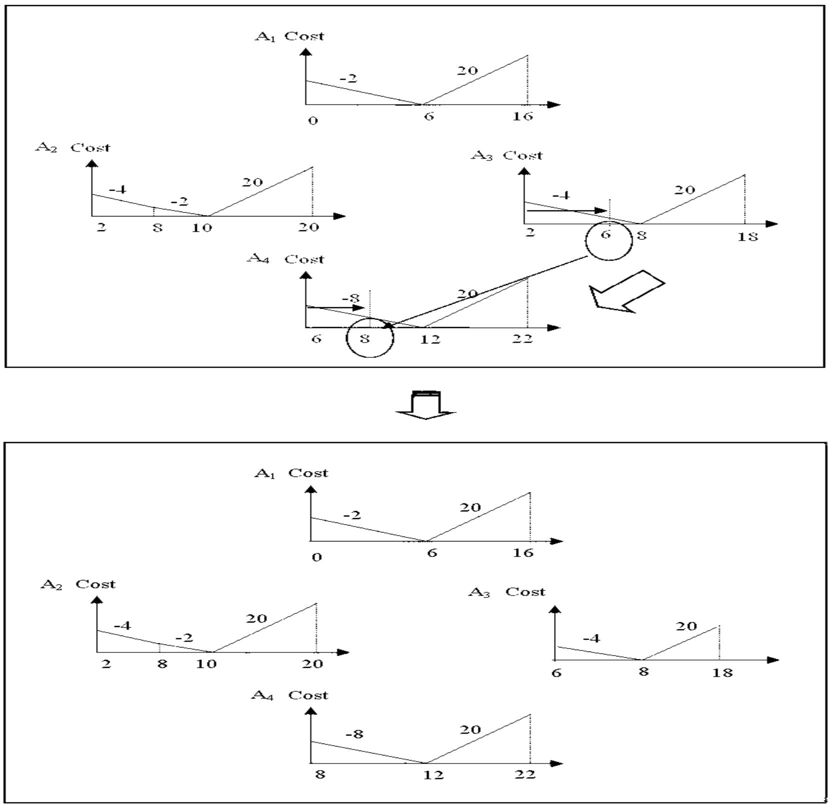
Upstream Propagation : If the optimal start time of activity A does not change, i.e., $est'_A \leq ost_A$, then there is no cost propagation for the activities upstream of activity A. When activity A has a new earliest start time greater than its optimal start time as shown in [Figure 8], the optimal start time of activity A increases too and cost propagation



[Figure 8] When activity A_3 increases its earliest start time from 2 to 10, the tardiness CC of its upstream activity A_1 is converted into an inventory CC for the interval [6, 8], and in turn, the slope of the inventory CC in activity $A2$ decreases from -2 to -4 for the time interval [8, 10].

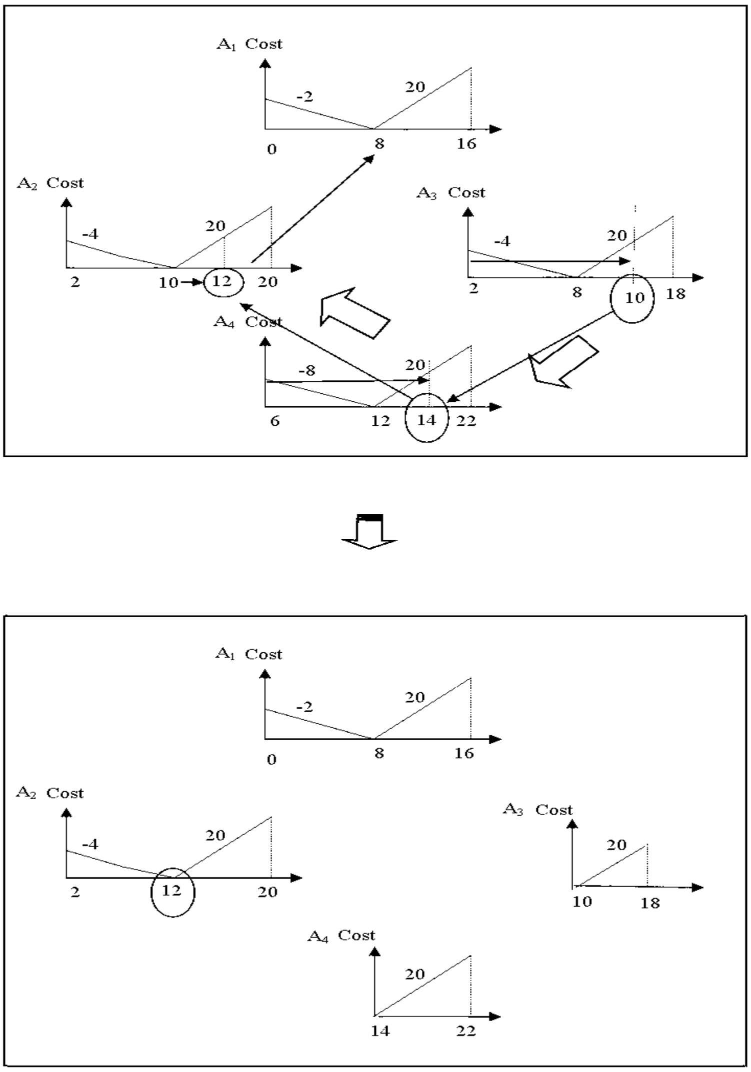
goes up to the upstream activities. In [Figure 8], activity A_3 causes a tardiness cost for the interval [8,10]. Hence, the upstream activity A_1 should move its optimal start time from 6 to 8. There is no change in the feasible time window of activity A_1, but the tardiness CC for the interval[6,8] should be converted into an inventory CC due to the increase of the optimal start time. The expansion of the inventory CC in activity A_1 affects the slopes of inventory CCs in its downstream activities on the non-critical path. Look at the new inventory CC of activity A_2 in [Figure 8]. The slope of inventory CC for the time interval[8,10] in activity A_2 is decreased from -2 to -4 since the time gap between the optimal start times of activities, A_1 and A_2 disappears.

Downstream Propagation : By temporal propagation, the earliest start time of each activity *A-f* direct downstream of activity *A* may be increased to $est'_{A\text{-}f}$ where $est'_{A\text{-}f}$ =Max(est'_A + du_A, $est_{A\text{-}f}$). Only the domain of CCs for *A-f* shrinks without no slope change. [Figure 9] shows that when activity A_3 increases its earliest start time from 2 to 6, its downstream activity A_1 simply increases its earliest start time as a result of temporal pro-



[Figure 9] When activity A_3 increases its earliest start time from 2 to 10, the downstream temporal propagation increases the earliest start time of activity A_4 from 6 to 8.

pagation. In a case that the optimal start time of activity *A-f* changes due to temporal propagation, the upstream cost propagation described above goes up, if exist, to the activities upstream of activity *A-f* as shown in [Figure 10] where the tardiness CC of activity A_2 is converted into an inventory CC for the time interval[10,12].



[Figure 10] When activity A_3 increases its earliest start time from 2 to 10, the optimal start time of activity A_4 increases to 14, and in turn, the tardiness CC of activity A_2 for the time interval [10,12] is converted into an inventory CC.

5. Concluding Remarks and Future Research Directions

The central aim of this paper is to present intra-job CPAs for propagating cost information globally across temporal constraints. The suggested algorithm is analyzed as follows. Updating inventory and tardiness CCs in the job having n number of activities requires at most *O(n)* time in each search state, when there is no activity on the non-critical path. In case that k number of activities reside on the non-critical path, the worst-case complexity becomes *O(kn)*. The space complexity is *O(n)* as we maintain individual CCs for each activity.

Based on the intra-job CPA, cost propagation is extended into other jobs connected through precedence relationships, by the inter-job CPA. By utilizing the cost information provided by intra-job CPA as well as inter-job CPA, it is possible to propose cost-based scheduling heuristics that attempt to minimize the total schedule cost.

So we plan to present inter-job CPA, and suggest cost-based constraint-directed scheduling heuristics.

References

[1] Beck, J.C., Davenport, A.J., Sitarski, E.M., and Fox, M.S., Texture-based heuristics for scheduling revisited, Proceedings of AAAI-97, 1997a.

[2] Beck, J.C., Davenport, A.J., Sitarski, E.M., and Fox, M.S., Beyond contention : extending texture-based scheduling heuristics, Proceedings of AAAI-97, 1997b.

[3] Caseau, Y. and Laburthe, F., Improved CLP scheduling with task intervals, Proceedings of the 11-th international conference on logic programming, 1994.

[4] Easton, F. and Goodale, J., Schedule Recovery : Unplanned Absences in Service Operations, Decision Sciences, Vol. 36, No. 3(2005), 459~488.

[5] Fox, M.S., Constraint-directed search : a case study of job-shop scheduling, Ph.D. thesis, Carnegie Mellon University, 1983.

[6] Le Pape, C., Implementation of resource constraints in ILOG Schedule : a library for the development of constraint-based scheduling systems, Intelligent Systems Engineering, Vol. 3, No. 2 (1994), 55~66.

[7] Nuijten, W.P.M., Time and resource constrained scheduling : a constraint satisfaction approach, Ph.D. thesis, Eindhoven University of Technology, 1994.

[8] Sadeh, N., Looking-ahead techniques for micro-opportunistic job shop scheduling. Ph.D. thesis, Carnegie Mellon University, 1991.

[9] Sadeh, N., and Fox, M.S., Variable and value ordering heuristics for the job shop scheduling constraint satisfaction problem, Artificial Intelligence, Vol. 86, Np. 1(1996).

[10] Smith, S.F., and Cheng, C.C., Slack-based heuristics for constraint satisfaction scheduling, Proceedings AAAI-93(1993), 139~144.

[11] Sun, W., Shi, W., Shi, B., and Yu, Y., A cost-effieient scheduling algorithm of on-demand broadcasts, Wireless Networks, Vol. 9, No. 3(2003).

요약

비용기반 스케쥴링 : Part I, 작업내 비용 전파알고리즘

김재경* · 서민수**

문제의 제약조건을 명확히 표현하고 휴리스틱 탐색에 의하여 스케쥴링을 형성하는 제약조건 중심의 스케쥴링 기법은 실세계의 스케쥴링 문제에 성공적으로 적용되어 왔다. 하지만, 기존의 제약조건 중심의 스케쥴링 연구에서 스케쥴링의 목적을 표현하고 최적화하는데 관련된 연구는 부족한 상황이다. 특히 비용 목적함수는 다양한 비즈니스 계획의 효과를 분석하는 기업의사결정에서 매우 중요하다고 평가된다. 이 연구의 목적은 자원 용량이나 일시적인 제약조건을 만족하면서 지연비용 및 재고비용을 포함한 스케쥴링의 전체 비용을 명확하게 표현하고 최적화하는 것이다. 비용기반 스케쥴링 프레임워크에서, 동일한 작업 내에 일시적인 제약조건을 만들어 가면서 비용함수를 개선해 나가는 비용 전파 알고리즘을 제시하였다.

* 경희대학교 경영대학
** 현대제철