Schedule Uncertainty Analysis System Framework to Manage and Allocate Historical Data for Industrial Construction Project

I. Yoon[a], H.-S. Lee[a], M. Park[a], J.G. Lee[a], and S.-S. Jung[a]

[a]Department of Architecture and Architectural Engineering, Seoul National University, South Korea
E-mail: yoon92411@snu.ac.kr, hyunslee@snu.ac.kr, mspark@snu.ac.kr, potgus@snu.ac.kr, tjs6191@snu.ac.kr

Abstract –

Many project managers and researchers have emphasized the importance of considering uncertainty in the scheduling of EPC projects, and perform stochastic analysis when preparing a schedule to reflect uncertainties of duration for project activities. In this case, the uncertainty of the individual activity duration is treated as probabilistic variables, and this probability should be estimated based on previous project experience and data. To use past project data for probabilistic schedule analysis of planned schedules, the schedule information of past project should be stored and managed systematically, and the analysis should be based on the relevant past project data. Existing schedule analysis tools are focused on the analysis process, and there is a lack of consideration on how to manage and apply the past project data needed for the schedule analysis. Therefore, this paper proposes a system framework that enables probabilistic schedule analysis, which automatically supports the process of identifying and allocating corresponding past project activities for estimating activity durations of a planned schedule. We propose a data model and system framework that reflects defined issues and the requirements. The framework is composed of 4 modules: As-built schedule data storage module, a planned schedule input module, a historical data allocation module, and uncertainty analysis module. It is expected that the proposed system framework will contribute to overcoming the difficulties of data management for collection and planned schedule analysis in practice.

Keywords –

Schedule Analysis; Probabilistic Analysis; Uncertainty; System Framework; Data Model

1 Introduction

In the construction industry, many mega-projects are delivered in the form of EPC contracts. In the early stages of the EPC project development, the project scope is unclear, because the EPC contractors are responsible for the design, procurement, and construction phases. This implies the uncertainty of the management factor such as cost and schedule is much larger than other construction projects. In particular, scheduling is critical for successful project management when considering the size, participants, and scope of EPC project [1,2], and the scheduling for initial project planning should be done with careful consideration of the uncertainty.

Many project managers and researchers have emphasized the importance of uncertainty considerations in the scheduling of EPC projects [3,4], and perform stochastic analysis when preparing a schedule to reflect uncertainties of duration for project activities. For scheduling of EPC project schedule, a planned schedule is prepared, which consists of hundreds, or even thousands, of activities. In this case, the uncertainty of the individual activity duration is treated as probabilistic variables, and this probability should be estimated based on previous project experience and data [5,6].

To use past project data for probabilistic schedule analysis of planned schedules, the schedule information of past project should be stored and managed systematically, and the analysis should be based on the relevant past project data. The reliability of the schedule analysis depends on the appropriateness of the utilized data and should select relevant information of past project activities should be selected to corresponding activities of the planned schedule. In this process, it is difficult to allocate the appropriate past information on the planned schedule due to the unclear project scope at the initial stage of the project, the difficulty of activity identification from a past project, inconsistency in resource and productivity to be put into each activity of EPC projects. Such problems indicate that the process of applying past project data to the probabilistic schedule analysis is a process that requires multiple decision-making, and it is impossible to allocate the past data to all activities of planned schedule in the same way.

To support the process of managing past project data and probabilistic schedule analysis, various methodologies and software have been developed and utilized for project scheduling. However, existing tools are focused on the analysis process, and there is a lack of consideration on how to manage and apply the past project data needed for the schedule analysis.

Therefore, this paper proposes a system framework that enables probabilistic schedule analysis, which automatically supports the process of identifying and allocating corresponding past project activities for estimating activity durations of a planned schedule. First, we investigated the precedent methods and studies for probabilistic schedule analysis to define challenging issues of project schedule analysis and the system requirements. And we propose a data model and system framework that reflects defined issues and the requirements.

2 Literature Review

In this chapter, the authors review uncertainty analysis technique to identify required data structure of the system and previous approaches about data structure for schedule data management.

2.1 Required data structure for uncertainty information

To deal with uncertainty of activity duration quantitatively, a probabilistic approach was generally used for schedule analysis. Monte-Carlo Simulation (MCS) is the most popular tool for probabilistic analysis. In case of scheduling, MCS is possible to be performed by considering activity duration as a random variable. During each iteration in the MCS process, values of random variables are determined from a specific range, probability distribution in other words [7]. Therefore, the reliability of result from uncertainty analysis depends on how to estimate the probability distribution of each activity.

According to previous studies and project management body of knowledge (PMBOK), which is recognized as a fundamental rule of project management around the world, if historical data exists sufficiently, high accuracy duration estimation is possible [5]. However, most of the researches [6,8–10] apply probability distribution to activity by expert judgment like three-point distribution due to lack of data availability. Therefore, for reliable results of analysis, it is an important issue how to develop data structure which is able to enhance data availability.

2.2 Previous Efforts for Improving Data Availability

Generally, a schedule includes work breakdown structure (WBS) information which has a hierarchical structure. Therefore, how to develop standard WBS has been an important topic in the area of construction data management. The approach for developing standard structure of WBS was based on the expectation that it could allocate the past performance information to the planned schedule of the new project if a company manages all projects with single standard WBS. Although the hierarchical structure has advantages in terms of management in practice, it has disadvantages such as difficulty in representing various perspectives and levels of detail [11]. To overcome these limitations, facet-classification model with facet concept emerged [12], and work package model was also proposed, which is improved based on the facet-classification model. Work package model expresses the structure of information divided into Where, What, How, and Who. Cho et al. [11] proposed a 5W1H model by improving abovementioned concepts.

These efforts can be categorized as a project-

Table 1. Comparisons of project-oriented data structure and activity-oriented data structure

Contents	Previous Approach	Proposed Approach
Concept	Project-oriented data structure	Activity-oriented data structure
Description	If the schedules based on a standard data structure, duration data can be allocated to new schedule just by collecting past project schedules.	Add past performance to the database through separate tasks after project execution
Advantage	If the standard is appropriate, no extra effort is required to collect and allocate schedule data.	-Flexible configuration of the purpose and type of data -Little effort is required to accumulate new data after initial database construction.
Disadvantage	If the characteristics of the each project are significantly different, there is difficulty in establishing the standard.	-The allocation process is cumbersome.

oriented data structure. In other words, previous researches try to establish a data model which is possible to utilize when a project schedule is created based on the fixed data structure. However, it is hard to ensure applicability in project schedule management practice with the project-oriented concept. Due to the different characteristics of every individual project, especially in case of mega-project, a more detailed breakdown of WBS or work package makes it hard to establish a standard data structure for multiple projects. To address the issues, this research proposes activity-oriented data structure to manage and to allocate historical data efficiently. Table 1 summarizes advantage and disadvantage of both approaches.

3 Data Model

Data for probabilistic schedule analysis can be acquired from past project schedule which contains actual data. However, data collection for certain purpose cannot be performed without purpose-built designed data model. The fact that the project manager has some past schedule files does not mean he or she can analyze schedule with historical data. So systemically structured database is essential to manage and allocate historical data to the planned schedule. A data model is an abstract model to represent data relationship and flow so that the reader can understand actual data better. The main objective of the data model is to enhance understanding of data structure and to provide access to the appropriate source to improve accessibility [13,14].

Fig. 1 shows developed logical entity relationship diagram of the proposed data model to represent activity-oriented data structure. Project and activity table in the figure indicate as-built schedules information. Resource table does not mean the type of resource, but the resource that has performed the actual project. Therefore, even if the same welder has two different activities, it would have two different resource ID. Information in the resource table is limited to labour in this research.

The duration information and the productivity information table on the right side of the figure indicate the storage type of data with a specific purpose. For example, if the A/G filed pipeline welding is to accumulate data on productivity, it is generated by assigning one data code to the Duration Information table. Whenever the information in the data-resource table is accumulated one by one, the values of productivity information such as the number of data, mean or max value of data are continuously changed. The tables in blue dotted line on figure represents the link of accumulated data to activity or resource information. These two tables are filled with information by the user’s direct information allocation. However, when the work package information operates to allocate other data and activities, the query is configured so that no data is inserted to facilitate later retrieval.

Also, activity, resource and each information table have attributes to the work package. The name of these attributes may vary depending on the user or project. This is because the way and perspective of managing projects differ by person, project, or company. Through this database development, it directly supports the as-built schedule data storage module and the historical data allocation module of the system framework.

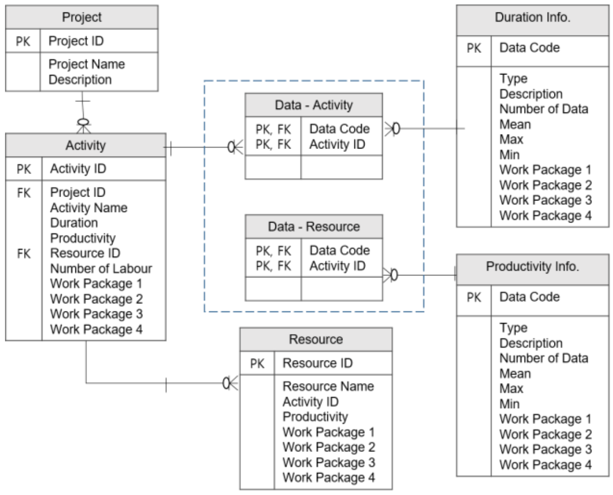

Figure 1. Proposed Data Model

4 System Framework

This chapter introduces the system framework to manage historical schedule data and to analyze the planned schedule with stored data. Each content of the system framework will be described respectively with modules and use case of the system. The prerequisite for using the system is that both as-built schedules and the planned schedule should use the same work package. The work package, which is meant here, is a high-level facet, such as a discipline (e.g. civil, architecture, mechanics) and phase (e.g. engineering, procurement, construction), so that even if the characteristics of the project are different, it can be shared by various projects. The work package is the basis for enhancing the usability of the storage and allocation modules to be described in the following chapters. Fig 2 shows the system framework developed in this research and each module of the system.

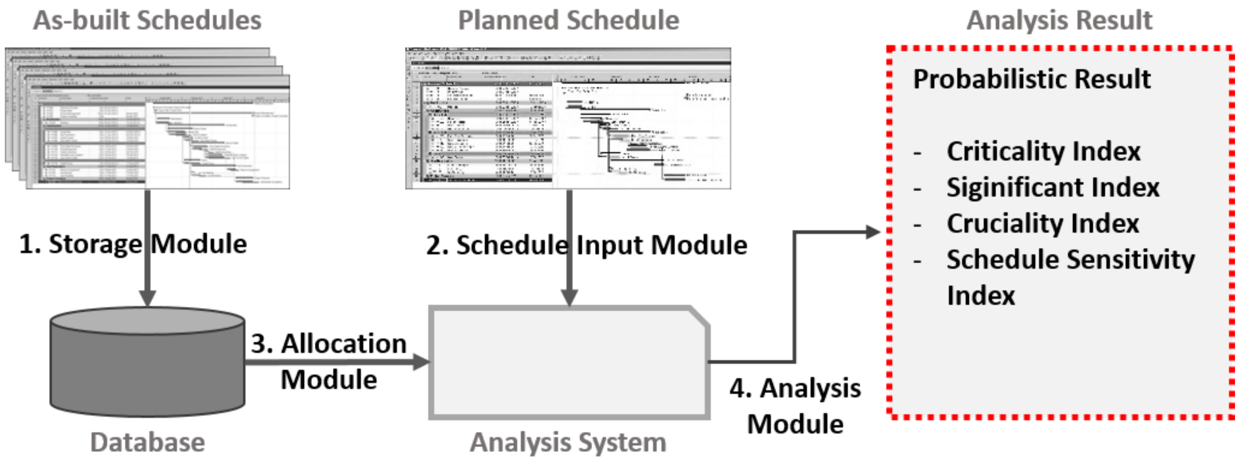

Figure 2. System framework for probabilistic schedule analysis

4.1 As-built Schedule Data Storage Module

After the project finished, the project manager has to store as-built schedule in the system. Storage module supports this process based on the data model developed in the previous chapter. In this module, the user stores the tasks of the as-built schedule in the database. In this process, the related data codes are retrieved through the work package and are selected. At this point, the user needs to decide whether to save the activity in duration table or to save in productivity table. For example, some activity or procurement tasks in engineering are often difficult to estimate for reliable productivity. Therefore, the reliability of the database can be improved by making such a selection directly by the user.

4.2 Planned Scheduled Input Module

In general, the EPC project schedule is developed and managed in Primavera or MS project. Therefore, to enhance a practical application, the system has to include a module that enables a connection with the existing schedule management system. Planned schedules generally include the following four information: activity, dependency, resource, and resource allocation. Both Primavera and MS project provides an ability to export this information as excel format. And we can easily convert excel to CSV (comma-separated variables) format. Based on imported these 4 CSV files, the system creates a simulation model of the project. This simulation model generation automation process is a modification of research in Jung et al. [15].

4.3 Historical Data Allocation Module

Although the planned schedule is input in the system, it doesn’t mean appropriate historical data is allocated to each activity in the planned schedule. Therefore, users have to retrieve relevant historical data in as-built schedules. Utilizing the attributes of the work package of the previously defined data model, users can find the data codes associated with each activity of the planned schedule. Data codes can be used to search for past related activities and to use the entire or some data inherent in the data code for the analysis.

After allocating historical data for activities of the planned schedule, the system is ready for MCS. It means that system can generate a probability distribution from historical data. From statistics information of historical data, the system can generate cumulative distribution(CDF) function, inverse cumulative distribution function(ICDF) in order. By entering a random number in the range between 0 and 1 into ICDF, a value according to the probability distribution can be obtained.

4.4 Uncertainty Analysis Module

Through the process described in the first three chapters mentioned above, the planned schedule entered into the system is ready to be analyzed based on historical data. When MCS is performed, virtual schedules are generated by the number of times of simulation. The statistics of these schedules can be used to create an indicator of how the activity's uncertainty affects the schedule planned schedule. There are following four popular indicators in previous researches: criticality index, significant index, cruciality index and schedule sensitivity index [16–18]. Based on the theoretical concept and the system framework, it is possible to provide these indicators to the user as simulation results.

5 Conclusions

This research proposed a system framework that supports the process of storing information on past schedules in a database and linking them with planned schedules. Also, this study proposed a required data model for the proposed approach. It is expected that the proposed system framework will contribute to overcoming the difficulties of data management for planned schedule analysis in practice.

There are still following limitations in the proposed framework. Most of them depend on heuristic tasks in the storage module and an allocation module. For the usability of the proposed system framework, future study is needed to explore the possibility of automation and implementation. Also, since the system is not yet implemented, the applicability and effectiveness of the proposed system framework should be validated after developing and testing a system prototype.

Acknowledgements

This work was supported by the Technology Innovation Program (10077606, To Develop an Intelligent Integrated Management Support System for Engineering Project) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

References

[1] Gepp M., Hellmuth A., Schäffler T. and Vollmar, J. Success factors of plant engineering projects. *Procedia Engineering*, 69:361–369, 2014.

[2] Yeo K.T. and Ning J.H. Integrating supply chain and critical chain concepts in engineer-procure-construct (EPC) projects. *International Journal of Project Management*, 20(4):253–262, 2002

[3] Alsakini W., Wikström K., and Kiiras J. Proactive schedule management of industrial turnkey projects in developing countries. *International Journal of Project Management*. 22(1): 75–85, 2004.

[4] Wang P. and AbouRizk S.M. Large-scale simulation modeling system for industrial construction. *Canadian Journal of Civil Engineering*, 36(9):1517–1529, 2009.

[5] PMI. PMBoK Guide: A Guide to the Project Management Body of Knowledge. 2013.

[6] Tokdemir O.B., Erol H., and Dikmen I. Delay Risk Assessment of Repetitive Construction Projects Using Line-of-Balance Scheduling and Monte Carlo Simulation. *Journal of Construction Engineering and Management*, 145(2): 04018132, 2019

[7] Goodarzi E., Ziaei M., and Lee T.S. Monte Carlo Simulation. Introduction to Risk and Uncertainty in Hydrosystem Engineering. Springer, Dordrecht, 2004.

[8] Schexnayder C., Knutson K., and Fente J. Describing a Beta Probability Distribution Function for Construction Simulation. *Journal of Construction Engineering and Management*, 131(2):221–229, 2005.

[9] Barraza G.A. Probabilistic Estimation and Allocation of Project Time Contingency. *Journal of Construction Engineering and Management,* 137(4): 259–265, 2011.

[10] Choudhry R.M., Aslam M.A., Hinze J.W., and Arain F.M. Cost and Schedule Risk Analysis of Bridge Construction in Pakistan: Establishing Risk Guidelines. *Journal of Construction Engineering and Management*, 140(7): 04014020, 2014.

[11] Cho, D., Russell, J.S., and Choi, J. Database Framework for Cost, Schedule, and Performance Data Integration. *Journal of Computing in Civil Engineering*, 27(6): 719–731, 2013.

[12] Kang L.S. and Paulson B.C. Adaptability of Information Classification Systems for Civil Works. *Journal of Construction Engineering and Management*, 123(4):419–426, 1997.

[13] Steeve Hoberman. Data modeling made simple. Technics Publications, Bradley Beach, NJ, 2005.

[14] Ji S., Park M., Lee H., Ahn J., Kim N., and Son B. Military Facility Cost Estimation System Using Case-Based Reasoning in Korea. *Journal of Computing in Civil Engineering*, 25(3):218–231, 2010.

[15] Jung M., Park M., Lee H.S., and Chi S. Multimethod Supply Chain Simulation Model for High-Rise Building Construction Projects. *Journal of Computing in Civil Engineering*, 32 (3):1–12, 2018.

[16] Vanhoucke M. Using activity sensitivity and network topology information to monitor project time performance. *Omega,* 38(5):359–370, 2010.

[17] Vanhoucke M. On the dynamic use of project performance and schedule risk information during project tracking. *Omega*, 39(4):416–426, 2011.

[18] Elmaghraby S.E. On criticality and sensitivity in activity networks. *European Journal of Operational Research*, 127(2):220–238, 2000.