Interactive Uncertainty Analysis

**Pedro Szekely, Rajiv T. Maheswaran, Yu-Han
Chang, Yan Wang, Huihui Cheng
and Karan Singh**

Information Sciences Institute
University of Southern California
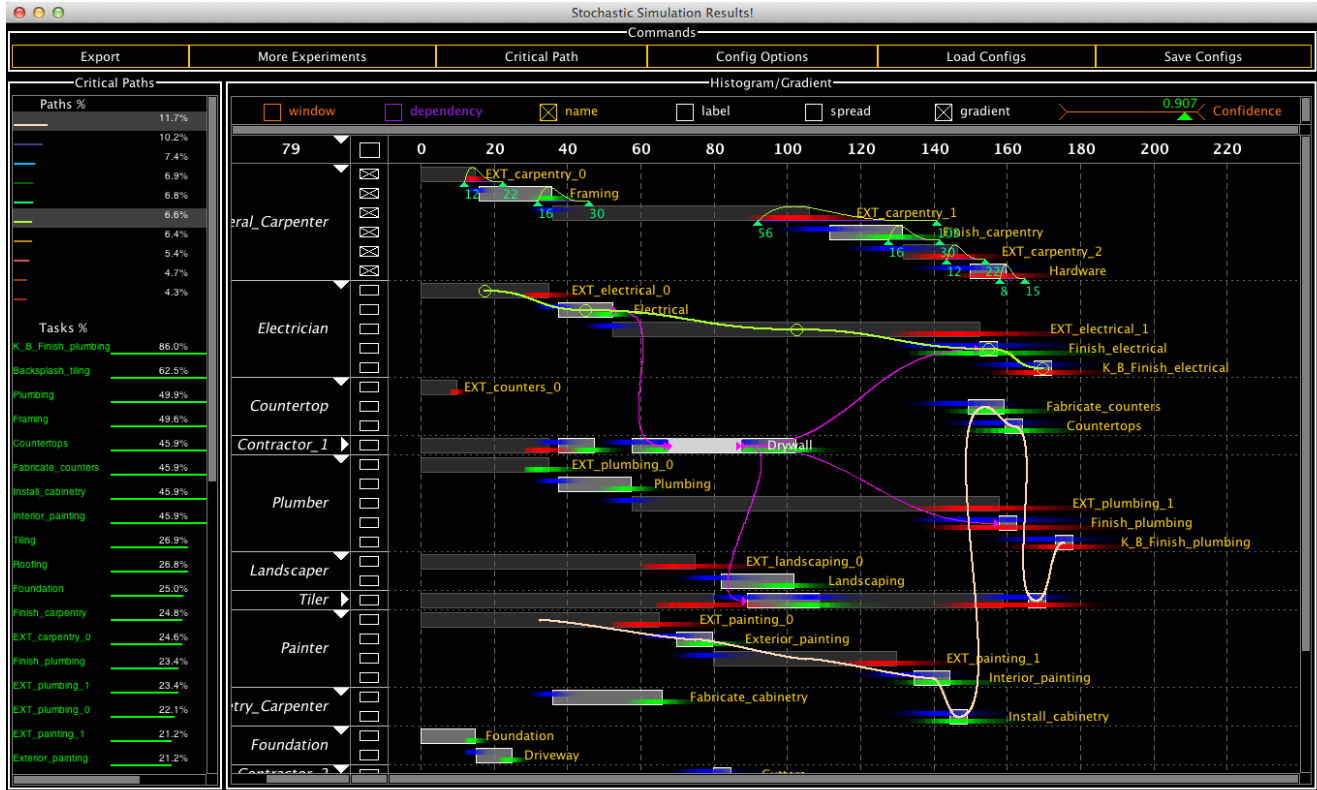{pszekely, maheswar, ychang}@isi.edu
{wang79, huihuich, karans}@usc.edu

Figure 1. Visualization of schedule uncertainty analysis using COMPASS.

ABSTRACT

Humans have difficulty evaluating the effects of uncertainty on schedules. People often mitigate the effects of uncertainty by adding slack based on experience and non-stochastic analyses such as the critical path method (CPM). This is costly as it leads to longer than necessary schedules, and can be ineffective without a clear understanding of where slack is needed. COMPASS is an interactive real-time tool that analyzes schedule uncertainty for a stochastic task network. An important feature is that it concurrently calculates stochastic critical paths and critical tasks. COMPASS visualizes this information on top of a traditional Gantt view, giving users insight into how delays caused by uncertain durations propagate down the schedule. Evaluations with 10 users show that users can use COMPASS to answer a variety of questions about the possible evolutions of a schedule (e.g., what is the likelihood that all activities will complete before a given date?)

Author Keywords

Visualization, Uncertainty Analysis, Plan Understanding, Schedule Visualization, Gantt Chart, Monte Carlo Simulation.

ACM Classification Keywords

I.6.6 Simulation and Modeling: Simulation Output Analysis; I.6.6 Simulation and Modeling: Types of Simulation—*Monte Carlo*

General Terms

Human Factors

1

INTRODUCTION

Uncertainty is inherent when scheduling tasks but unfortunately humans are poor at calculating its effects. The problem is exacerbated in task networks of non-trivial size as nonlinear effects (e.g., when a task depends on the completion of multiple tasks) cascade. Given a task network, a user will have multiple questions of different types that they want answered. This can range from questions about the given plan, such as when it's expected to finish or the likelihood of making a deadline, to what-if scenarios (e.g., what happens if it rains? what if everything takes two days longer than anticipated?). Our goal is to support real-time interactions to answer these questions to help users get an overall understanding of how uncertainty affects their plans.

While others have worked on various aspects of the problem (see Related Work), none have provided real-time what-if capabilities over multiple aspects of the plan, a visual representation of uncertainty and its effects on the entire plan, nor a validation through user studies. Our contribution is COMPASS (see Figure 1), a tool which (1) performs stochastic analysis of schedules, (2) enables real-time interactivity and feedback, (3) provides a comprehensive and intuitive visualization of the entire plan, and (4) includes diverse types of what-if capabilities. In addition, we (5) conducted a user study that shows significant improvements in the accuracy of and confidence in the participants' answers involving stochastic schedules.

INTERACTIVE UNCERTAINTY ANALYSIS

We describe COMPASS and its capabilities in the context of a user study we performed to evaluate its effectiveness. Users were given a plan to build a house involving 40 tasks (partially shown in Figure 1) with 11 agents and 30 enablement dependencies. They were then asked to answer questions such as determining the expected plan completion time, 90% confidence interval for the completion times, and what-if questions that involved changing the uncertainty distributions. We modeled task durations using beta distributions ($\beta(2,10)$). With these skewed bell-curves, durations range from 80% to 200% of the mean, and 99% of the durations are between 80% and 140% of the mean. Users can see and modify the distributions by selecting the tasks in the display.

To help users answer these questions, COMPASS performs thousands of Monte Carlo simulations of schedule execution (typically between 2 and 10 thousand simulations) satisfying all constraints in the temporal network and using a execution policy where agents perform tasks as soon as they can (agent is free and all task constraints are satisfied). The COMPASS visualizations consist of layers that summarize the information gathered in these simulations.

When will it be done? We asked the users to answer deeper variants of the question "when will a given task be complete?":

1. What is the expected time when the house will be finished?
2. At what day will we be 90% confident that the house will be done? (i.e., 90% of the time, the house will be finished before this day).
3. What is the probability that our house will be finished within 180 days?
4. How wide is the uncertainty window for completing a task if we want to be 90% confident that the task completes within that window?

Users can answer the questions above using the *window* and *gradient* layers, and the *confidence* slider (see Figure 2). The time windows enable users to specify earliest and latest desired times for task completion, and are shown as thin rust color lines behind the task boxes. The gradients layer displays histograms constructed from the Monte Carlo simulations of the schedule execution. The histograms for each task are visualized as color gradients on the time line, one for the start time and one for the end time. The intensity of the color at each time point is proportional to the number of simulation runs that started and ended at the given time point. The gradients for start-times are blue, and the gradients for end-times are green or red depending on the time windows and the value V of the confidence slider (between 0 and 1): If the fraction of the area of an end-time histogram that falls within the time window is larger than V, then the gradient is green, otherwise it is red.

In Figure 2, the left panel shows the histograms at 80% confidence: finish_carpentry and EXT_Carpentry_2 are expected to complete within their time windows with 80% confidence, while the other two tasks are not. When users drag the confidence slider, the colors update dynamically, enabling users to see the interplay between confidence and the time windows, seeing histograms change to red as they increase the desired confidence threshold. In Figure 2), the right panel shows the state when the user drags the confidence slider until the histogram for EXT_Carpentry_2 turns red, indicating that there is only a 90% confidence that this task will complete within the desired time window.

The user can answer the first two questions by manipulating the confidence slider, and moving the mouse to appropriate positions on the visualization and reading the time marker. The third question can be answered by defining the time windows for tasks and then moving the confidence slider until the corresponding histogram changes color (see Figure 2). For the fourth question, users can turn on the *spread* layer, which shows the spread window for the start and end-time of all tasks. Users can manipulate the confidence slider and see how the spreads change as a function of it. In schedules that use beta distributions, spreads increase dramatically as confidence approaches one.

To help users understand how uncertainty propagates, when users move the mouse over a task, COMPASS draws purple arrows from the enabling tasks to the enabled tasks. Figure 1 shows the enablement lines for the Drywall task in the middle

Figure 2. Time window, gradients and confidence sliders to analyze the uncertainty of task completion.

of the screen.

Which are the most critical tasks? In the traditional, non-stochastic Critical Path Method, a critical path consists of the set of tasks with the property that if any of these tasks are delayed, it will delay the completion of the schedule. Here the answer to the question is simple: the most critical tasks are those in the critical path, since any delay in these tasks will cause the schedule to slip (their slack is zero). If task durations are uncertain, the traditional definition is not appropriate because tasks can have a range of durations, and it is not clear which one should be used in the computation of the critical path.

To answer the question in the model with uncertainty, COMPASS computes the critical path of each simulation run, and then builds a histogram of all critical paths. Figure 1 shows the visualization of critical paths. The left panel lists the different critical paths and the percentage of traces where they occur. Users can compare critical paths visually by selecting them and seeing the curves that connect the tasks in each path. In our example, the most common critical path occurs 11.7% of the time, involving tasks for the painter, carpenter, counter-top installer and the plumber. Below the list of critical tasks is a list of tasks sorted by the likelihood that they lie on some critical path. This helps users understand the criticality of tasks and focuses their attention on the tasks that are more likely to cause schedule delays. For example, even though the most critical path occurs only 11.7% of the time, the most critical task is part of 86% of all critical paths.

What will happen if . . . ? Many interesting what-if questions can be expressed in terms of changes to the duration distributions of sets of tasks. For example:

1. If it rains, and every task affected by rain becomes twice as uncertain (e.g., the worst-case time doubles), when will we be 90% certain that our house is done?
2. If we believe the electrician underestimates his duration estimates by two days, when will our house be completed on average if we shift all his estimates upwards by two days?

COMPASS offers several ways to select tasks and alter distributions. Users can select tasks individually by clicking on them, and they can select all the tasks performed by an agent by clicking on the agent. A more sophisticated selection mechanism uses the *label* layer, where users can see and associate arbitrary labels (keywords) with tasks. Users can select all tasks with a given label by clicking on any of the labels. When tasks are selected, their duration distributions appear. Users can simultaneously modify all the distributions for the selected tasks by editing the distribution of any one of them. They can shift the distribution in either direction, and all other distributions are shifted by the same amount. They can drag the minimum or maximum time while keeping the other constant, and other distributions are scaled by the same amount. Figure 3 shows an example where the labels layer was activated, and the user selected all tasks labeled rain.

Using COMPASS, users can answer the first question by selecting all the tasks labeled rain, and then dragging the right edge of any of the distributions so that the distribution becomes twice as wide. When users release the mouse COMPASS re-runs the Monte Carlo simulations and within 2 seconds updates the display, showing how the new distributions affect the schedule. Similarly, users can answer the second question by clicking on the electrician to select all its tasks, dragging them two days to the right, and seeing the effect on the whole schedule.

Together, these features enable users to visualize how uncertainty affects the schedule, to investigate how external events that affect multiple tasks affect the whole schedule (e.g., rain), and to investigate the effects of alternative resources with different characteristics (e.g., faster, more reliable). Video: http://youtu.be/vmT86VXTN9Y

EVALUATION
Our evaluation, conducted with ten subjects, consisted of three 30-minute parts. In part 1, we familiarized users with the house schedule by giving them hardcopies of the Gantt chart and task dependency graphs, and explaining the schedule to them. In part 2, we trained users on COMPASS using a small sample plan consisting of 3 agents, 5 tasks, and 4 enablement dependencies. We introduced the COMPASS features one at a time, and asked users to practice it on the sample schedule before introducing the next feature. In part 3, we asked users to answer "when", "criticality" and "what-if" questions similar to those shown in the previous sections. We gave users 30 minutes to answer 20 questions.

Subjects answered 81% of the questions correctly. If we omit data from two users who had trouble with the direct manipulation techniques, the top 8 users answered 91% of the questions correctly. Based on exit surveys, the other major source of error was the users' general misunderstanding of the notion of spread.
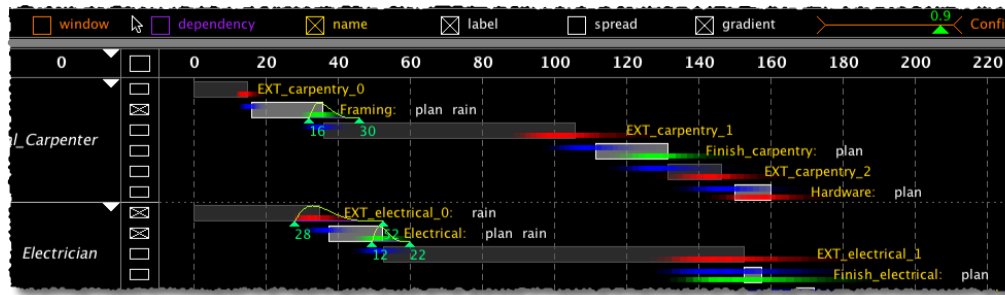
Figure 3. Tasks labeled *rain* are selected; their duration distributions can be simultaneously edited.

RELATED WORK

Clark and Gantt [4] published their seminal work on Gantt charts in 1923. This early work is noteworthy not only because it introduced Gantt charts, but also because it describes several variants of the basic Gantt chart that use additional symbols and annotations to encode additional information for decision making. While they did not address the issue of uncertainty, our work on embellishing Gantt charts to encode information useful for understanding uncertainty, follows the same approach. Our work is most similar to the work on Planning Lines [1] that also embellishes the basic Gantt charts to encode uncertainty by using shades of color to denote different intervals of uncertainty. Planning Lines focuses on conveying information about minimum and maximum duration times and earliest and latest start and completion times, and thus uses only a fixed number of colors. Tory et al. [6] also augment Gantt charts to show task delays and aid with schedule comparison. Our work quantifies the uncertainty using histograms, which we encode using color gradients.

Work on understanding uncertainty in schedules also has a long tradition. PERT [5] focused on "most likely time", "optimistic time", and "pessimistic time". More recent work [3, 2] focused on quantifying uncertainty using Monte Carlo techniques. This work focused on the mechanics of valid stochastic simulations, without paying heed to the need for user comprehension and interaction. For example, Lee [8] and Lana [7] introduced software that performs stochastic simulation and simple criticality analysis. After providing the plan parameters and assumptions through dialog boxes, the user is simply given the resulting distribution of plan completion times, without any insight about where the delays begin, how they propagate, where slack is helpful, or where to focus efforts to improve completion confidence. Our work visualizes the results of the simulations. However, the main difference between the previous work and our work is that the COMPASS visualizations are active. We use the dynamic query techniques first introduced by Shneiderman [9] to enable users to ask questions of the visualization, and thereby ask what-if questions about schedules.

CONCLUSIONS

The user study suggests that users understand the COMPASS visual language and can use it to answer questions regarding uncertainty. A surprising aspect of this work is that informal demonstrations to engineers and researchers revealed that they are unable to answer the questions in the user study without the system. COMPASS made them realize they had ignored an important issue in their reasoning. Our initial user study revealed that some users had trouble manipulating the small graphic elements of our visualization. We are working to refine the visualization and the interaction techniques, and on techniques to compare schedules so users can better understand the before and after states of their what-if scenarios.

REFERENCES

1. Aigner, W., Miksch, S., Muller, W., Schumann, H., and Tominski, C. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics 14* (2008), 47–60.

2. Bowman, R. A. Efficient estimation of arc criticalities in stochastic activity networks, 1995.

3. Burt, J. M., and Garman, M. B. Conditional monte carlo: A simulation technique for stochastic network analysis. *Management Science 18*, 3 (1971), 207–217.

4. Clark, W., and Gantt, H. L. *TThe Gantt chart, a working tool of management*. New York, Ronald Press, 1923.

5. Fazar, W. Program evaluation and review technique. *The American Statistician 13*, 2 (1959), 646–669.

6. Huang, D., Tory, M., Staub-French, S., and Pottinger, R. Visualization techniques for schedule comparison. *Computer Graphics Forum 28*, 3 (2009), 951–958.

7. Lana, E. H. Stochastic critical path. In *Crystal Ball User Conference* (2006).

8. Lee, D.-E. Probability of project completion using stochastic project scheduling simulation. *Journal of construction engineering and management 131*, 3 (March 2005), 310–318.

9. Schneiderman, B. Dynamic queries for visual information seeking. *IEEE Software 11* (1994), 70–77.