QUALITATIVE STUDY OF USER ANNOTATIONS IN SATELLITE SCHEDULING

by

Susan Villecroze
B.Sc., University of Victoria, 2001

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the School of Computing Science

© Susan Villecroze 2007

SIMON FRASER UNIVERSITY
Summer 2007

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.
APPROVAL

Name: Susan Villegroze
Degree: Master of Science
Title of Thesis: Qualitative Study of User Annotations In Satellite Scheduling.

Examining Committee:
Chair: Greg Mori
Assistant Professor

______________________________
Arthur Kirkpatrick
Senior Supervisor
Assistant Professor

______________________________
M. Stella Atkins
Supervisor
Professor

______________________________
William Havens
Internal Examiner
Associate Professor

Date Defended/Approved: July 26, 2007
Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the Library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website <http://ir.lib.sfu.ca> at: <http://ir.lib.sfu.ca/handle/1892/112>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada

Revised: Summer 2007
STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada
ABSTRACT

Past research in scheduling has focused on algorithmic issues and has not addressed many important human-computer interaction issues. For tasks that require a higher level of abstraction and decision, annotation tools could provide an aid. This study investigated how people used annotations to solve problems presented on printed schedules. A user study involving 5 participants was conducted. Participants were presented with a pre-computed satellite schedule and given a practical problem to solve. Video observations, interview answers, and markings on the schedule and source documents provided data for analysis. Results show that while making trade-offs on different priorities, every participant used and benefited from the use of annotations. Participants did not always use specific annotations because of the connotations of the annotation appearance. The results suggest that support is needed for marking priority changes, deleted activities, interesting regions, and adding text on the schedule.

Keywords: mixed-initiative interaction; satellites; scheduling; annotations; tagging; annotation behavior; user study

Subject Terms: Human-computer interaction -- Psychological aspects; User interfaces (Computer systems); Remote sensing -- Congresses; Photographic surveying
DEDICATION

To my loving husband Jean-Louis, who has supported me throughout this entire venture.
ACKNOWLEDGEMENTS

I would like to thank my senior supervisor, Ted Kirkpatrick. His insightful criticism and guidance are what has led me through the completion of this project. He has supported me throughout my graduate career and is a role model when it comes to teaching.

Thank-you Sarah Brown for implementing the satellite scheduler. It was not used for this study but I did get to use it for my first pilot studies. After studying the interface and trying to fix bugs in the software, I find myself respecting the work and the people that build schedulers even more.

I must thank James Peltier, the Technical Director in the GRUVI lab, for helping me with the equipment. I felt I could always depend on him and in each time of need, he was quick to respond. James surely is the “know-it-all” in the lab.

Thanks go out to the other members of my examining committee: Stella Atkins and Bill Havens. Thank-you Stella and Bill for taking time out of your busy schedules to read my paper and to come to my defence. Your helpful comments were much appreciated.

I also have a couple of personal acknowledgements to make. I am grateful to my parents for raising me and providing me with great work ethics. It was not easy for them but I know that they did their best. My success in the future would also be their success.

Lastly, and most importantly, I owe thanks to my husband Jean-Louis for his love and encouragement. He has tolerated me through the years, and even more so through my difficult journey at SFU. He provided me with emotional support, laughter, and a purpose in life. He is my best friend. Without him, I would not be in grad school in the first place.
# TABLE OF CONTENTS

- Approval .................................................................................................................. ii
- Abstract .................................................................................................................... iii
- Dedication ................................................................................................................ iv
- Acknowledgements ................................................................................................... v
- Table of Contents ....................................................................................................... vi
- List of Figures ........................................................................................................... viii
- List of Tables ............................................................................................................. ix
  1. Introduction ........................................................................................................... 1
    1.1 EOS Scheduling Problem .................................................................................. 3
  2. Related Work .......................................................................................................... 5
  3. Background and Objective ................................................................................. 8
    3.1 Project Background ......................................................................................... 8
    3.2 The Paper Study and Its Objective .................................................................. 9
    3.3 Proposed Outcomes and Hypotheses ............................................................... 11
  4. Pre-computed Sample Schedule ......................................................................... 12
  5. User Study ............................................................................................................ 13
    5.1 Participants ...................................................................................................... 13
    5.2 Equipment and Setup ...................................................................................... 14
    5.3 Procedure ........................................................................................................ 14
    5.4 Experiment Task .............................................................................................. 16
    5.5 Strengths and Limitations of User Study ......................................................... 16
  6. Results ................................................................................................................... 18
    6.1 Observations during the Task .......................................................................... 18
    6.2 User 3 Observations ....................................................................................... 19
    6.3 Participant Markings on the Schedule .............................................................. 20
    6.4 Participant Markings on Source Documents .................................................... 23
    6.5 Interviews ........................................................................................................ 26
    6.6 Why Annotations Were Used on the Schedule ............................................. 30
    6.7 Classifying the Annotations Found ................................................................. 31
    6.8 Participant Feedback ....................................................................................... 33
  7. Discussion ............................................................................................................. 34
    7.1 Implications for the Design of Mixed-Initiative Schedulers ......................... 34
    7.2 Limitations of the Experimental Protocol ....................................................... 35
    7.3 Markings and Tools Used ................................................................................ 37
    7.4 Comments about the Task .............................................................................. 38
    7.5 Comments on User 1 ....................................................................................... 39
    7.6 Problems and Issues ....................................................................................... 40
  8. Directions for Future Work ................................................................................... 41
  9. Conclusions .......................................................................................................... 43
- Reference List .......................................................................................................... 45
LIST OF FIGURES

Figure 1: Schedules displayed in the scheduler .......................................................... 10
Figure 2: Printed schedule used in the user study ..................................................... 12
Figure 3: First ten minutes of the printed schedule ................................................. 13
Figure 4: Video snapshot displaying the view from the camcorder ............................ 14
Figure 5: The section of User 1’s schedule near the new activity ............................ 21
Figure 6: The section of User 2’s schedule near the new activity ............................ 21
Figure 7: The section of User 3’s schedule near the new activity ............................ 22
Figure 8: The section of User 4’s schedule near the new activity ............................ 22
Figure 9: The section of User 5’s schedule near the new activity ............................ 23
Figure 10: Some markings on the front of User 1’s task sheet ............................... 24
Figure 11: Some markings on the back of User 1’s task sheet ............................... 24
Figure 12: Activity IDs listed on User 2’s task sheet .............................................. 25
Figure 13: Some markings on User 2’s reference sheet ......................................... 25
Figure 14: Some markings on User 5’s reference sheet ......................................... 25
Figure 15: Activity #10 was marked as having higher priority .............................. 27
Figure 16: Association on User 2’s task sheet ......................................................... 31
Figure 17: Underlined numbers on User 1’s task sheet ......................................... 32
Figure 18: Emphasis used on User 5’s task sheet ................................................... 32
LIST OF TABLES

Table 1: Summary of how annotations were used on the schedule ........................................30
Table 2: Summary of annotations found for each participant .................................................33
1. INTRODUCTION

There are many scientific and commercial Earth Observing Satellites (EOS) in orbit around the Earth capturing images of the planet’s surface. EOS scheduling is an oversubscribed scheduling problem; that is, there are far more requests than can possibly be satisfied. To satisfy as many observation requests as possible, operators use scheduling programs to help plan the image-taking activities of the satellites. Satellite scheduling is a hard problem to solve; in fact, it is NP-hard in computational complexity. Constraints for the problem include fixed image locations, imager slew and duty cycle, power, thermal, data capacity, downlink locations, and the limited time each satellite spends over each target (Globus, Crawford, Lohn, & Pryor, 2004). EOS also need to respond to unexpected meteorological conditions such as cloud cover, targets of opportunity such as natural disasters, and changes in satellite or ground station capability (Khatib, 2004). It is possible for some satellites to make hundreds of observations per day and there could easily be thousands of request backlogs.

By involving users, optimization systems such as satellite schedulers can leverage people’s abilities in areas in which they outperform computers, such as visual and strategic thinking (Anderson et al., 2000). Dividing the labor into those areas where the computer excels, such as extensive search through alternative possibilities to find a desired solution, and those where human input is valuable is known as mixed-initiative planning and scheduling (Kramer & Smith, 2002). Mixed-initiative refers to a flexible interaction strategy, where each agent (human or computer) can contribute to the task what it does best. A mixed-initiative approach to scheduling is also important since it allows users to gain trust in the system. Users will remain reluctant to use the system unless they can question the system’s decisions and perform what-if analyses to explore solutions (Smith, Lassila, & Becker, 1996).

There are other reasons why constant human supervision and control is crucial. Algorithms hardly ever succeed in capturing all aspects of the real world. Satellite scheduling problems, like
all scheduling problems, are dynamic. This may be due to poor estimates, incomplete data, or unanticipated disturbances. Most previous automated systems failed to support users' tasks or problem solving due to a common lack of attention to the user (Cortellessa, Cesta, Oddi, & Policella, 2004). In all practical cases, the user wants to be kept in charge of the decisions taken by the tool. In satellite scheduling, users of interactive optimization systems can further improve solutions by manipulating the systems while respecting constraints. As a result, users can better understand, implement, justify, and modify generated solutions.

As a possible aid to further help users understand plans and develop better ones, annotation tools could be introduced to an interactive satellite scheduler. People are known to annotate to improve comprehension of source material or for later review, to assist in working out problems, to record interpretative activity, and to share work with others. The most commonly-cited benefits are to help recall and support search (Marlow, Naaman, Boyd, & Davis, 2006). Furthermore, navigation in an environment without distinctive visual appearance is difficult (Lewis, Rosenholtz, Fong, & Neumann, 2004). Trafton and Trickett (2001) found that people who take notes while actively engaged with a problem-solving task, perform better than those that do not. In their study, it appeared that note taking boosted simple recall of learned material, helped students comprehend data they have gathered, and possibly helped them to develop good problem-solving strategies. Although taking notes may be helpful for students during the early stages of their problem solving efforts, when they are still trying to comprehend the task and the steps necessary to solve it, Trafton and Trickett (2001) state that it may not be as important after students understand what they are doing.

Due to the complexity in scheduling satellites, many studies have compared algorithms for improved performance, but few use interactive systems that make use of annotations. ROSE is one of the few schedulers that features annotation tools (Wilkinson et al., 1995). In fact, interactive scheduling systems have not been studied in detail. There have also been many
studies on annotation tools, but only a few of them have examined user annotation behavior. Of those studies that looked at behavior, most of them (Marshall, 1998; Wolfe, 2000) examine text on printed documents in one way or another.

Cognitive mapping is a process by which an individual acquires, codes, stores, recalls, and decodes information about the relative locations and attributes of phenomena in their everyday spatial environment (Downs & Stea, 1973). Cognitive maps are most effectively formed by active interaction with the environment and an inaccurate or incomplete one leads to confusion. Marking regions with annotations could provide the visual cues needed to build memory and develop cognitive maps while performing a complex task. In this paper, we investigate how annotations are used and present our findings in a study using printed schedules for Earth observing satellites (EOS). To our knowledge, this is the first study that examines users’ annotation behavior in the context of scheduling.

The rest of this report is organized as follows. Some satellite scheduling terminology is introduced when EOS scheduling is described in more detail in the next section. This is followed by a short description of the related research to date. Then, the project’s background history, the objective of the study, and the proposed outcomes and hypotheses are given. The pre-computed schedule used for the study is then described. This section is then followed by a detailed description of the user study and a report of the findings. Finally, a summary is presented and the results are used to discuss design implications for future work.

1.1 EOS Scheduling Problem

Satellite scheduling is the problem of mapping activities (observations, communications, downlinks, control maneuvers, etc.) to resources (satellites, sensors, ground stations, etc.) (Pemberton & Galiber, 2000). The problem is also known as a single-resource scheduling problem where a set of activities competes for the use of a single resource (Baruah, Gehrke, & Plaxton, 1996). Only one of them is able to use the resource at a time, the resource is scheduled
in unit-time intervals, each activity requires part of the resource capacity over an extended period, and activities arrive and leave at any moment. Scheduling involves assigning one or more resources and a start time to each activity in a way that satisfies the constraints on the activities and resources. An example activity could be the following (Pemberton & Galiber, 2000):

*Satellite S must downlink to ground station G for 2 minutes between 23:05 and 23:27.*

In EOS scheduling, images are taken of specified locations on the Earth surface, in response to observation requests. Each observation or activity is assigned a request ID and may be scheduled at several points in the day. The visibility of the satellite to its target is used to define event windows consisting of start and stop times for a given activity-resource pair. In the above example, the event window is from 23:05 to 23:27. Data gathered and stored onboard the satellite is to be transferred to a ground station on Earth during downlink connections.

A scheduling program is required to satisfy as many observation requests as possible, favoring those with higher priority. In most practical domains, scheduling is viewed as an iterative process. The scheduler pre-computes a plan, which essentially tells the satellite what to do at any given moment. An operator then recognizes problems, relaxes and tightens constraints and requirements, and revisits the solution. Under this mixed-initiative user control, users can continually interact with the system at different levels of automation. For instance, when the removal of an activity is required in the schedule, a “frozen out” command can be issued to the scheduler. Activities can also be “frozen in” to guarantee that they will be scheduled in all future updated plans.

The user’s activity in a general scheduling task can be classified into three categories (Chimani, Lesh, Mitzenmacher, Sidner, & Tanaka, 2005): inspection, modification, and user controlled re-optimization of the current solution. During the inspection stage, the user is the one in control. He or she can explore the solution and check to see if constraints are satisfied. The application provides the appropriate visualizations, both for giving an overview, and to examine
specific details. If the user is not satisfied with the current solution, he or she can move into the modification stage and make changes to the schedule such as requesting the scheduler to keep or remove activities. Re-optimization can occur after an inspection or modification stage. Here, the intent may be to concentrate computational resources on optimizing an important part of the schedule, to fix an infeasibility introduced during modification, or simply to re-optimize some portion of the schedule after making a change (Chimani et al., 2005).

2. RELATED WORK

Scott, Lesh, and Klau (2002) evaluated several individual components of human “in-the-loop” optimization. Their paper described experiments on an interactive optimization system that explored the most appropriate way to combine the respective strengths of people and computers. Several user tasks within a “human-in-the-loop” optimization system were examined and users’ performance in these tasks was compared to the performance of the same task by the computer. A tabletop display was used and a HuGSS\(^1\) system was studied since it allowed users to guide optimization of capacititated-vehicle-routing-with-time-windows (CVRTW\(^2\)) problems. Participants repeatedly performed various subtasks of the overall optimization process, which included a focusing task, finding-targets task, and stopping task. They found that users could successfully identify promising areas of the search space as well as manage the amount of computational effort expended on different subproblems.

Obendorf (2003) compared the support for active reading using traditional paper and pencil to a tablet PC using two of the existing annotation tools for the Web: Annozilla and Webnize Highlighter. Paper was found to be a much better reading device than the tablet PC with either

\(^1\) The human-guided simple search (HuGSS) framework allows users to manually modify solutions and steer the optimization process itself.

\(^2\) The CVRTW problem is a variant of the traveling-salesman problem where the objective is to find the least-cost solution based on the number of trucks and the distance they drive that satisfies all the constraints.
annotation tool. It was clear that participants tried to transfer their annotation practices from paper over to the computer.

Ramos and Balakrishnan (2003) developed a system called LEAN that serves as an exploratory platform for a variety of visualization and interaction techniques for navigation, segmentation, linking, and annotation of digital videos. LEAN runs on pressure-sensitive digitizer tablets. One of the goals of the system is to allow users to navigate and annotate digital video with fluidity and ease similar to that used on printed documents using pens and post-it notes. Ramos and Balakrishnan’s study also researched techniques for annotating video. For instance, LEAN allowed for a note to be “pinned” into the workspace using a discrete pressure widget, which would then make it visible at all times. In a preliminary study, six users were asked to explore the system freely and were encouraged to engage in tasks that involved navigating and annotating a video clip. The results indicated that being able to freely annotate and link items in a workspace can be advantageous. The exact advantages were not given.

Fogli, Fressta, and Mussio (2004) introduced a set of tools for web document annotation for a two-way exchange of ideas among humans pursuing a common goal. Their paper discussed the importance of “e-annotation” as a tool by illustrating a scenario where a photo-interpreter and a glaciologist are collaborating in order to arrive at the classification of a glacier.

Fu, Ciszek, Marchionini, and Solomon (2005) conducted a survey study on making annotations for personal use when viewing web pages. Their goal was to understand Web users’ needs for annotation tools to make the kinds of annotations one sees on paper. Text selection and emphasis, building association, and re-segmentation of the document were examined in the Web environment since these were the most common forms of annotations on printed documents. The dominant forms of annotation on the Web were found to be text selection and association building through notes. Re-segmentation and layout annotation such as change of font or color were also popular. The findings also indicated that users welcomed easy-to-use, lightweight
annotation functions built into standard Web browsers. Unfortunately, by their admission, there were a few problems with their study and they fell short in answering which factors affect personal annotation behavior. For instance, their results had limited generalizability due to the study's exploratory nature. In addition, the participants were not representative of the Web user population and only a few had previously used all the annotation tools in the study.

Collaborative tagging and annotation systems are a growing area of interest in the human-computer interaction (HCI) community. Sen et al. (2006) presented a user-centric model of vocabulary evolution in tagging communities based on community influence and personal tendency. Their model was evaluated in the MovieLens recommender system, after introducing tagging features into the system. They analyzed the effect on vocabulary evolution, tag utility, tag adoption, and user satisfaction from four tag selection algorithms. They examined factors that influenced both the way people choose tags and the degree to which community members share a vocabulary. With regards to a user's personal tendency, it was found that habit and investment influence users' tag applications, habit and investment influence grows stronger as users apply more tags, and habit and investment cannot be the only factors that contribute to vocabulary evolution.

MacKay (2000) studied the use of paper flight strips in air traffic control. Like satellite scheduling, air traffic control is a complex, collaborative activity and the work requires rapid responses to constantly changing conditions. The differences are that little to no scheduling is performed, safety is the primary concern, and controllers rely on their ability to extract information from the environment, as they need it. Paper flight strips allow air traffic controllers to organize the traffic, plan their strategies, modify information about planes and flight plans, and record key questions. A four-month ethnographic study was undertaken at the Athis Mons (Paris, France) en-route control center, as well as a comparison study of eight air traffic control centers.

---

3 MovieLens is a free service provided by GroupLens Research at the University of Minnesota. It is a movie recommendation website located at http://movielens.umn.edu.
in France and the Netherlands. The controllers’ annotation practices on the flight strips were described as part of the results. Annotations were written because they were required and they helped controllers develop a mental representation of the air traffic. Controllers annotated strips in predictable ways that were easily interpreted by other controllers. They learned specific rules for how to annotate strips, but individual styles as to how much to write and when varied greatly. For example, student controllers wrote the most to help learn and to demonstrate that they knew the rules. It was also found that different controllers indicated the same thing with different annotations and different teams had different opinions about different types of annotation. Also, controllers were more likely to annotate strips in medium traffic levels than in stressed situations. In stressed situations, controllers spoke less often and reduced writing to a minimum. MacKay concluded with implications for the design of safety-critical systems. One alternative to using paper flight strips is to keep the current system, which is extremely safe and efficient, and just augment it.

3. BACKGROUND AND OBJECTIVE

3.1 Project Background

In 2005, Brown and Kirkpatrick started a satellite scheduling project that aimed to look at annotation behavior while performing all three categories of the user’s activity: inspection, modification, and re-optimization. They built a scheduler that implemented ideas in many academic papers (Kramer & Smith, 2002; Cortellessa et al., 2004; Cesta, Cortellessa, Oddi, & Policella, 2004). An annotation feature was then added to the system. Annotations in the form of images could be added over activities that are scheduled. Operators could also “freeze-in” or “freeze-out” activities before updating a plan or allow the system to update the plan. New plans could also be created and plans could be deleted. Experiment documents, including a detailed study protocol, were also written but no user study had been done using the scheduler.
In November 2006, the first pilot study was conducted using the scheduler developed by Brown and Kirkpatrick (2005). Two graduate students in computer science participated in the study and none of them had prior experience working with satellite scheduling. The exact study protocol that was outlined by Brown and Kirkpatrick (2005) was carried out and proved to be too difficult for the participants. There were problems with the software that needed to be fixed. For instance, developing a real scheduler that accounts for power level is complicated and was not done properly.

The user annotation study in this research attempts to continue the work by Brown and Kirkpatrick (2005). After many attempts to use the existing scheduler, the decision was made to abandon it completely. Even with a simpler task, pilot studies still proved hard for participants. No annotations were used from the program and the participant’s approach to solving the problem did not match the behavior of satellite schedulers that we wished to model. Due to time constraints, not all of the software problems could be resolved.

We decided to take a step back and perform a paper study with the schedules. In fact, Brown and Kirkpatrick (2005) conducted an informal paper study prior to developing the scheduler. Their experimental task resulted in participants making little to no annotations. This led them to believe that their paper task was not a good model since users were not engaged in the iterative process of making a schedule. Consequently, they developed a simulated satellite scheduler. In the early stages of this project, we attempted to design tasks and fix bugs in the scheduler to accommodate annotation behavior. Unfortunately our efforts failed. We concluded that conducting another paper study is the most efficient and cost-effective way to accomplish this study.

3.2 The Paper Study and Its Objective

In this study, we have designed a hard but feasible task and provided the participants with a printed schedule that is based on what the current scheduler displays (Figure 1). The schedule is
explained in detail later. Here, the user’s annotation behavior is only examined during the inspection stage. Unlike the previous paper study, this study presents a scenario in mixed-initiative interaction and participants are going to find it very hard to complete the task without making any marks. The way in which people annotate on paper provides insight into how users may annotate on a computer (O’Hara & Sellen, 1997). We believe this paper study is an important step towards designing a proper experiment and suggesting improvements for the simulated satellite scheduler.

Figure 1: Schedules displayed in the scheduler

We are interested in how the use of annotations on a printed schedule can improve a user’s understanding. By understanding, we mean that users are able to use annotations to help them solve a problem. We address the following questions:
• How do people use annotations (if at all) as they work to understand printed plans that are given to them? How does annotation behavior compare between participants?

• Do people use specific annotations because of the connotations of the annotation appearance or are arbitrary marks used? How are meanings assigned to the marks that are used?

• What recommendations can be made for the design of the scheduler implemented by Brown and Kirkpatrick (2005)?

Although we were interested in how people used annotations to help them understand schedules, we did not attempt to make any statements about what people were thinking. Rather than focus on understanding cognition, we aimed to provide accurate descriptions of what was performed.

3.3 Proposed Outcomes and Hypotheses

Based on the literature review and our pilot work, we believed that annotations would be made on the given plans. In particular, participants would at the very least use annotations to mark places they would like to return to, like a bookmark. When presented with a complex problem with various factors to consider, we believed it would be very difficult to solve everything entirely from memory and it would only be natural to outline, select objects, and make notes on paper. After all, paper and pen provides the appropriate affordances for this sort of active engagement (Marshall & Brush, 2004).

We also believed that in most cases, people would use specific annotations because of their connotations. By using marks in this way, people could easily remember what had already been done and that would reduce their cognitive load.
4. PRE-COMPUTED SAMPLE SCHEDULE

A pre-computed schedule was used in the user study. The values for this schedule such as the length of the activity and the power level were computed using Microsoft Excel and a small script. The scheduler by Brown and Kirkpatrick (2005) was then modified to accept the values as a file input. Four different sets of values were computed, resulting in four different plans.

Figure 2: Printed schedule used in the user study

Screenshots of Plan 2 were taken and then modified and pieced together in Photoshop (Figure 2). This schedule was then printed on a single sheet of paper of size 6x55 inches. The first ten minutes of the schedule is shown in Figure 3. There are two panels on the schedule: the top displays the activities being scheduled and the bottom shows a graph of the corresponding battery power. Boxes represent the images or scheduled activities. For simplicity, only low-priority images are scheduled in the plan. The numbers on the boxes are the request IDs. The time scale on the schedule is in minutes and the power supply runs from zero to 180 power units. The text and numbers above an activity correspond to the length of the activity, measured in seconds, and the power required to schedule that activity, measured in power units. For instance, if \( L=10 \) and \( P=15 \), then the length of the activity below it is 10 seconds and 15 units of power is required to schedule it. The total value of the plan is displayed at the top. The length of time that is required to observe a target or schedule an activity and its associated priority determines its value.
5. USER STUDY

5.1 Participants

A user study with five participants was conducted. Participants were graduate students in computing science. All five were male and three were between the ages of 24 and 27. Two participants declined to state their age. Everyone had taken a large range of mathematic courses such as algebra, calculus, statistics, discrete math, computational complexity, and algorithms.

Four people stated that their mathematical abilities were strong and one claimed that it was good. None of them had prior experience working with satellite scheduling and none had participated in any of the previous studies. When asked what their energy level was like, four people stated that it was normal while one claimed to be tired. The session lasted around one hour. Participants were paid $10 CDN for their involvement.
5.2 Equipment and Setup

A Canon XL1s 3CCD digital video camcorder was used to record video and audio. Recording was set to 25 frames per second and was done in LP mode. Audio mode was set on 16-bit for the highest sound quality. 16:9 standard was used as the video aspect ratio. Five 60-minute digital video cassettes were used to collect the data during the user study. The recordings were then imported to the computer and compressed into a movie using iMovie HD 5.0.2 on a Power Mac G5 featuring Mac OS X 10.4.9. The video camcorder was placed to the right of the table and participant, approximately 15-20 centimeters away from the edge of the table. The view captured the entire schedule. Figure 4 shows a snapshot of one of the videos.

5.3 Procedure

The experiment was conducted in a quiet usability lab at Simon Fraser University, Burnaby, Canada, on June 22, 2007. A printout of the agenda for the session (Appendix A) was used to guide the experiment in order to ensure that everyone was given the same information. The experimenter remained close by during the session in case user assistance was required. Scrap paper, color pens, color highlighters, stick-ons, a ruler, and calculator were provided for making
annotations. A countdown timer was also provided so that participants could tell how much time was left.

All participants were required to complete the same problem during the session. Each participant was given a total of three documents plus the schedule. The documents were the satellite scheduling explanation sheet (Appendix D), reference sheet (Appendix F), and user task description (Appendix E). A session consisted of two phases: inspecting a schedule to solve a problem and explaining the answers that resulted from it. Each session was roughly broken down as follows:

1. Welcome, fill out consent form and background information form (5 min)
2. Explanation of the satellite schedule (10 min)
3. Task: Inspecting a schedule and solving the problem (30 min)
4. Break (5 min)
5. Interview questions and post-session questionnaire (20 min)

Total: 70 min

Each participant started by filling out the consent (Appendix B) and background information forms (Appendix C). Then they were asked to read a sheet describing satellite scheduling (Appendix D) and the reference sheet (Appendix F). Both the schedule and the reference sheet were explained.

When time expired for the task, the participant was allowed a break for up to five minutes before the interview would start. During the interview, the experimenter asked questions about how the problem was solved and how marks were used (Appendix G). Questions on the post-session questionnaire were also asked (Appendix H). This was done to ensure that all questions were answered properly.

Video and audio were recorded during the task and the interview. Rough notes were taken as answers during the interview.
5.4 Experiment Task

The participant was given a detailed problem and had 30 minutes to solve it. The problem required the participant to resolve three issues in the schedule (see Appendix E). The first issue was that a problem was encountered at a ground station and the scheduled downlink could no longer occur. The second issue was that a set of activities currently scheduled had now increased in value. Finally, a new request had been made for a high priority image. The user’s task was to recommend modifications to the schedule while still operating within specified constraints. For instance, the battery reserves had to be above a minimal level at all times, indicated by the dashed line at the 30 unit mark on the graph (Figure 3). Other considerations and constraints had to be taken into account and were listed such as memory limitation, activity priority, and total plan value. Participants were told that we were not concerned about accuracy and that rough answers were also acceptable. They were also told that they were free to write on anything that was given to them. For those who wanted to crunch a few numbers, the power drain and power gain rate was given. These numbers were provided since participants in our pilot studies requested them.

5.5 Strengths and Limitations of User Study

The following is a list of strengths in the user study. A paper study had the advantage of being doable within the time constraints. Pilot studies with the printed schedule also showed that people annotated in order to solve the problem. As an added benefit, participants were free to annotate however they wished on paper using the tools that were provided. In contrast, the scheduler only permitted images to be added above activities. Activity boxes were also colored in when “freeze in” or “freeze out” commands were issued.

The rest of this section lists the limitations in the user study. Annotations that people make may depend on the task that is given to them. In this study, the same task was performed by all the participants in order for data to be compared more easily. Unless we explicitly test for it in a
future study, we cannot be sure whether or not certain annotations used in this study were task dependent.

The number of participants that are used may also affect the final results. If common behavior cannot be seen among a small group of people, then more participants would likely be used. Different people from the same group tend to behave differently but we did not suspect them to be very different since everyone had a common task and goal. However, if annotation behavior turned out drastically different for each participant, that would also be an interesting result.

The annotations that people make may also depend on the type of people we recruit for the study. If satellite schedulers were available for recruitment, then their participation may have resulted in fewer and different annotations used. Due to their expertise, they may not need to make as many marks as a computing science grad student with no prior experience in scheduling. Satellite schedulers may also use annotations that are standard to scheduling, which others outside the field would not understand. This would also mean that the annotations used by a group of satellite schedulers would have many similarities.

Another limitation is that the analysis was done by hand. The videos were reviewed on the computer but any quantitative measure had to be determined manually. For instance, imagine having to count the number of times an object is circled. This could be a potential problem in terms of accuracy and would worsen as the number of participants increase. On paper, circles are imperfect and could be misidentified, whereas if a program were used to apply circles, there would be no mistake. In general, there was the risk of mistaking marks to carry meaning when they did not since we did not ask questions about every mark on the schedule or any of the sheets given out.
6. RESULTS

Every participant made marks during the task and claimed that their markings helped them solve the problem. This finding supports our hypothesis that annotations are made when presented with a complex problem. Marks were concentrated in similar areas across the schedules. Markings were also made on some of the task sheets, reference sheets, and scrap paper. A few additional marks were made during the interview by a couple of people. With the exception of one participant, few tools were used to make markings. In most cases, only a pen and calculator were used. One person thought the task was easy; another thought it was hard, and three people thought it was reasonable.

This results section is organized as follows. A general overview of the video observations is given, followed by a detailed account of User 3’s behavior during the task. Then, the markings on the schedule, source documents, and scrap paper are described. This is followed by the answers given during the interviews. The results is then summarized according to their usage and then classified. This section concludes with some of the feedback given by the participants.

6.1 Observations during the Task

Each video was viewed from start to end once and then revisited whenever there was a need to confirm an answer from the interview notes or comment on a marking that was seen. Some of the videos were also reviewed when there was uncertainty in some of the results. The interview portion of the videos was viewed more than the portion on the participants’ task performance.

Video recordings revealed similar problem solving processes for many participants. In each case, the participant started by reading the task and reference sheet carefully, sometimes spending as much as eight minutes just reading. Participants occasionally used a pen to point and guide them along the way. As reading took place, the schedule was referred to and in many instances; markings were also made. Throughout the experiment, the task sheet, reference sheet, and
schedule were constantly being referred to. The timer was also monitored and consequently, everyone completed the task in the 30 minutes that was assigned.

Of all the participants, I found User 3’s interaction with the schedule particularly interesting. He made use of more tools than any other person and his video often showed movement, even while he was thinking. As a result, I have chosen only to describe details of his behavior during the task.

6.2 User 3 Observations

User 3 was the only one that claimed the task was hard and whom I found interacted the most with the schedule. During the task, User 3 made use of the red and black pen, yellow and pink highlighter, ruler, calculator, and scrap paper. Scrap paper was used to list numbers and the calculator was used to sum them up. Rather than using stick-ons, User 3 chose to rip up pieces of paper off the scrap paper he computed numbers on. Often times, User 3 was heard talking out loud while he performed the task.

Less than one minute into the study, User 3 added yellow highlights over the list of activities that had increased in value on the schedule. He then immediately picked up the black pen and continued to read. Three minutes later, scrap paper was snatched and User 3 began ripping off small pieces of it. Reading then commenced and another tool was taken. User 3 read another part of the problem, examined the schedule, selected a tool, and then applied a mark. This cycle more or less repeated itself until the task was completed.

About half way through the session, selected activity lengths were written on scrap paper and then summed up with the calculator. More pieces of paper were ripped and then used to cover some of the activities on the schedule. More numbers were jotted down and the little pieces of paper were shifted around. User 3 leaned back in the chair and appeared to be thinking while fidgeting with a pen. Finally, a red pen was used to mark “|” above the pieces of paper and then
the pieces were removed. After examining the power level, red “-” marks were added through all the red “|” marks to create a “+” sign above some of the activities.

6.3 Participant Markings on the Schedule

Markings on the schedules consisted of slashes or Xs through activities, rectangles in between activities, marks over priority activities, scribbles, circles, arrows, and vertical lines. With the exception of User 4, everyone also drew in lines or dots to reflect changes in the power level. The following identifies every mark that was seen on the schedules.

User 1’s schedule consisted of forward slashes over activities, scribbled out slashes with a checkmark underneath activities, small squares underneath activities with a “2” inside the squares, a new activity drawn in that resembled the others, a small square underneath the new activity with a “3” inside the square, some text, and vertical lines (Figure 5). User 2’s schedule consisted of arrows pointing down over the top of selected activities, circled numbers, big Xs over activities, a new activity faintly drawn in, and dotted vertical lines (Figure 6). User 3’s schedule consisted of red “+” signs over the top of selected activities, yellow highlighted marks inside activity boxes, a red X over the cancelled downlink connection, and a black and a pink vertical line. There was also an unattached piece of scrap paper with the new request information written on it, which was placed in the region where it should be added (Figure 7). User 4’s schedule had checkmarks underneath activities, forward slashes over activities, a backward slash over the cancelled downlink, a bracket over the time where the new request could be added, an unfilled box to represent the new request, a filled in box, and vertical lines (Figure 8). The vertical line at time 1:35 resembled the pink line on User 3’s schedule. On User 4’s schedule, the word “Down” was written over the top of this line to represent when the downlink connection was going to return. User 5’s schedule contained checkmarks, circled short vertical bars above activities, boxes faintly drawn in between activities, horizontal arrows, an X over an activity, circled items, and vertical lines (Figure 9).
Figure 5: The section of User 1's schedule near the new activity.

Figure 6: The section of User 2's schedule near the new activity.
Figure 7: The section of User 3's schedule near the new activity.

Figure 8: The section of User 4's schedule near the new activity.
Figure 9: The section of User 5's schedule near the new activity.

With the exception of User 4, everyone drew in lines or dots to reflect changes in the power level. User 2 and 5 drew in their power lines starting at the point where the downlink connections had been removed (Figures 6 and 9, respectively). It appeared that power lines were first drawn near the new request before User 2 and 5 realized that they should have also taken the cancelled downlink connection into account. User 1 and 3 only considered the power level changes near the region where the new request would be added (Figures 5 and 7, respectively).

6.4 Participant Markings on Source Documents

Every participant made notes of some sort on either the source documents or scrap paper. User 1 underlined key items on the task and reference sheet such as times and activity IDs. Many notes were also made on his task sheet (Figures 10 and 11). User 2 underlined and noted items on the task and reference sheet that puzzled him. For instance, "what limitation on memory?"

Down a column on the right margin on the task sheet, User 2 listed each activity ID that had
increased in value (Figure 12). User 2 also made notes on his reference sheet (Figure 13). User 3 only summed up numbers on a piece of scrap paper. User 4 started to compute something on the bottom of his task sheet but then stopped. User 4 did not make any other marks besides those that can be seen on the schedule. User 5 circled words on the reference sheet that would help him determine the memory limitation. The bottom of User 5’s reference sheet also had markings (Figure 14).

Figure 10: Some markings on the front of User 1’s task sheet.

Figure 11: Some markings on the back of User 1’s task sheet.
after, you are notified that the following tasks are now worth twice as much as before: 10, 15, 24, 33, 37, and 43-46. On top of that, you just received a new request. It appears that the new request is worth 3 times that of a basic image.

New request: A forest fire has just started in region R and needs to be observed. Region R will be approached by the satellite and can be photographed between 1:28PM and 1:30PM. The observation should last at least 20 seconds, but no more than 30 seconds.

Your task is to recommend modifications to the schedule while still operating within specified constraints. See if the new requests can be added and if so, what needs to be done, in terms of deletions and additions to the schedule? Also, assume at the start of the schedule the memory

Figure 12: Activity IDs listed on User 2's task sheet.

Figure 13: Some markings on User 2's reference sheet.

Figure 14: Some markings on User 5's reference sheet.
6.5 Interviews

The interviews (Appendix G) indicated that although everyone had a slightly different answer, the downlink problem in the task was generally solved using similar reasoning. A couple of people also misunderstood part of the problem. Everyone noted places where they wanted to return to in the schedule and everyone assigned meanings to their marks. User 3 was the only one that claimed that some of his marks were for communication. Everyone used marks to highlight their priorities.

Every participant prioritized the pool of activities based on their value. The new request was worth the most so it was scheduled first. If they thought more power was required, only basic images were removed. Secondly, the original activities that had doubled in value were kept in the schedule. When it came to solving the downlink problem, various strategies were employed by different people but were more or less similar. Everyone knew that in order to meet the memory limitations, some activities would have to be deleted. They also knew that in the worst case the number of activities that needed to be deleted had to correspond to what the satellite was initially able to store onboard, had there not been any problems at all. All answers given for the downlink problem were found acceptable by the experimenter.

A few of the participants recommended changes that were not permitted on the schedule. For instance, User 1 initially slashed out both activities corresponding to the first downlink connection but then later scribbled the first one out and placed a checkmark underneath it. The problem stated that both of these activities corresponded to the cancelled downlink so they should have both been removed. User 3 based his recommendations on his assumption that the downlink connection could be scheduled again at a later time. Because the satellite would be overlooking a different location in its orbit, this is not possible. When notified about his error, User 3 suggested an alternative solution. Similarly, User 4 thought he could add a downlink connection activity
when the ground station returned but later realized that this was not possible. This explains why there was a scribbled out rectangle drawn in his schedule at time 1:35.

When asked whether any marks were used to note places they wanted to return to in the schedule, everyone pointed out various marks on their schedule as examples. User 3 was the only one to reply “no” to this question but then quickly added how he used pieces of paper to block out activities until a decision was made. User 3 explained how he was careful not to make too many marks on the schedule by using the pieces of paper. User 1 mentioned that he underlined text on his task and reference sheet. He also noted the circled regions and the note he wrote which read “+45 power” on the schedule (Figure 5). User 2 made markings to reflect changes in the power graph, noted important jobs, and circled above an activity that he noticed had the exact length and power requirement as the new request. User 4 used a bracket to mark the time slot where the new request could be inserted and he also marked the time where the downlink is expected to return. User 5 mentioned that he noted places where he wanted to make changes and called them critical points.

Figure 15: Activity #10 was marked as having higher priority.
Everyone assigned meanings to their markings. All the activities that had doubled in value were marked differently (Figure 15). User 1 indicated these activities by drawing a 2 inside a square under the activity, while User 4 used checkmarks. User 2 placed an upside down arrow above the activities but User 5 used circled vertical bars. User 3 simply highlighted them yellow for emphasis.

On User 1’s schedule, all the forward slashed activities indicated that they should be removed. All other activities should be kept in the schedule, including the ones with a checkmark attached. Checkmarks were placed underneath some of the activities because they were mistakenly slashed out to begin with. The new request was worth three times the value of a basic image and therefore, the square underneath it contained a 3 inside. Two sets of activities on either side of the event window for the new request were circled because they needed to be examined for power adjustments (Figure 5). User 1 boxed in the event window using vertical lines. When asked whether any of the meanings assigned to the marks changed over time, User 1 stated “no”. He then added that he made a few mistakes in removing some of the activities on the schedule but then corrected himself afterwards.

User 2 marked important activities with an upside down arrow. These activities included activities that were two and three times the value of a basic image. The first X over the downlink activities indicated that they were removed. The second X in the schedule was a typo. The activities with their “P” values circled were the ones that needed to be deleted. User 2 claimed that the meanings he assigned to his marks did not change.

The red “+” signs above the activities on User 3’s schedule indicated that they should be deleted. A red pen was used because User 3 associates the color red with edits. He commonly uses a black pen to work things out so this was used on the scrap paper and power graph. Like User 2, the big X over the downlink activities indicated that they had been removed. When asked
if his assigned meanings of marks changed over time, he mentioned that he initially used paper to store the state after canceling an activity but then switched over to use the calculator.

Like User 1, User 4 also used a slash through an activity to indicate that they should be removed. The items with a checkmark underneath it were high priority items and should be kept in the schedule. The meanings he assigned to his marks did not change.

User 5 circled vertical bars to mark the activities that had doubled in value. The new request was drawn in and “X3” was written above it to indicate that it had three times the value of a basic image (Figure 9). Vertical lines were drawn at the starting edge of a few of the activities down to the power graph in order to show the corresponding energy level (Figure 9).

When asked whether any of the marks made were done as a way to share their answers, User 1, 2, and 4 stated that all their marks were personal. The participants were told beforehand that they would be interviewed afterwards to explain their answers. User 3 said that the red “+” signs were made for communication but the rest were personal. User 5 was seen circling items while explaining his answers during the interview but it was unclear from the video whether or not he was circling existing marks or adding new marks as a way to communicate.

The last question asked was how the use of marks helped them make decisions on which activities to keep or remove. User 1 responded by saying that they made the priorities clear so that he would not remove ones that were considered high priority. User 2 said that the marks definitely helped but could not expand further. User 3 stated that his pieces of paper allowed him to use a greedy algorithm to decide on what to cancel without making too many marks on the schedule. He also claimed that his yellow highlights were used so that he would not have to constantly refer back to the reference sheets. User 4 said that the checkmarks helped him to remember that the activities above it should not be deleted. Similarly, User 5 said that the marks helped him memorize what the priorities were.
6.6 Why Annotations Were Used on the Schedule

Based on all the markings seen on the schedule and the explanations given in the interviews, a summary can be given for why marks were made (see Table 1). This summary helps to address our first research question of how people use annotations.

Table 1: Summary of how annotations were used on the schedule

<table>
<thead>
<tr>
<th>Annotation Purpose</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computed effects on power</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Marked priority of change</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Marked “interesting” area</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Marked “interesting” point</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>Corrected an existing annotation</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Drew new request</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Labeled inside new request</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Marked activity to be deleted</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tested solutions using pieces of paper</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Added a textual note</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Most annotations were associated to activities on the schedule, but there was also the need to mark specific points in time. With the exception of User 2, everyone marked an “interesting” point or area on the timeline. User 1, 4, and 5 marked an area where the new request could be inserted. User 1 also circled activities to examine on either side of this time slot. User 2 noticed an activity on the schedule containing the exact length requirement as the new request and he circled it as an interesting point. This point was associated with an activity rather than the timeline. Tiny marks on the power graph below this point were also seen. It appeared as though User 2 attempted to transfer the power drain graph at this point over to the new request. User 3...
and 4 both marked when the downlink would return as an interesting point. User 4 also placed a note, which read “Down” by that point. User 5 wrote “X3” above the new request and User 1 wrote “+45 power” on the graph. Everyone but User 3 made annotation errors on the schedule and had to correct himself. User 1, 2, and 3 labeled their new request. User 1 wrote “new” inside the box, User 2 wrote “R”, and User 3 wrote “F.”

6.7 Classifying the Annotations Found

Marshall (1998) identified how five common elements were realized in annotations on paper. Used textbooks were analyzed one-by-one and annotations were classified. These elements were associations, anchors, emphasis, re-segmentation, and types and categories. Many of these elements have been cited by other studies, including the one by Fu et al. (2005) on the Web environment. Three of these five elements were seen in this user study. The first one is associating an annotation with text span, which includes notes that refer to subparts of a document. This was seen on User 1’s schedule (Figure 5) and User 2’s task and reference sheets. Figure 16 shows an association found on User 2’s task sheet.

Summary of issues to take care of:

1. You will have to reconsider the schedule now that memory will NOT be empty after task 30

Figure 16: Association on User 2’s task sheet.

The second element was anchors. They are spanning marks, such as brackets, used to distinguish a region of interest. User 1, 4, and 5 used anchors. User 1 used vertical bars to isolate the region where the new request should be inserted. User 4 used a bracket and User 5 used arrows to indicate this same region. The use of anchors can be seen on Figures 5, 8, and 9 between the times of 1:28 and 1:30.

The final element from Marshall’s (1998) list of elements that appeared in these data was emphasis. Emphasis is the use of marks next to an element to make them salient and to indicate
importance. It also allows people to sort through and organize their annotations (Marshall, 1998). Every participant used emphasis. At the very least, marking the priorities on the schedule was for emphasis. User 1 also underlined text and numbers on the task and reference sheets. Figure 17 shows some underlined numbers on User 1’s task sheet. User 2 listed each activity ID separately in the margin of the task sheet (Figure 12). User 5 circled important text on his reference sheet (Figure 18).

schedule currently assumes the satellite’s memory storage is empty after task 30. Shortly after, you are notified that the following tasks are now worth twice as much as before: 10, 15, 24, 33, 37, and 43-46. On top of that, you just received a new request. It appears that the new request is worth 3 times that of a basic image.

New request: A forest fire has just started in region R and needs to be observed. Region R will be approached by the satellite and can be photographed between 1:28PM and 1:30PM. The observation should last at least 20 seconds, but no more than 30 seconds.

Figure 17: Underlined numbers on User 1’s task sheet.

how the number and length of the tasks could be associated with the

Figure 18: Emphasis used on User 5’s task sheet.

Sometimes annotations are assigned types based on color or some visual symbol and a “key” or legend is presented, as in the case when types and categories are used. Although User 3 used a yellow highlighter to indicate high priority activities, it was the only coloring used and therefore was used for emphasis rather than to distinguish types. Re-segmentation was not seen and was not allowed on the schedule since activities could not be shifted around. This was due to the requirement that activities had to be linked to target locations on Earth. Table 2 summarizes the annotations found for each user. Each table entry states where the particular element was seen.
Table 2: Summary of annotations found for each participant

<table>
<thead>
<tr>
<th></th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Associations</strong></td>
<td>Schedule</td>
<td>Task/reference sheet</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>Anchors</strong></td>
<td>Schedule</td>
<td>–</td>
<td>–</td>
<td>Schedule</td>
<td>Schedule</td>
</tr>
<tr>
<td><strong>Emphasis</strong></td>
<td>Schedule, task/reference sheet</td>
<td>Schedule, task sheet</td>
<td>Schedule</td>
<td>Schedule</td>
<td>Schedule, reference sheet</td>
</tr>
</tbody>
</table>

6.8 Participant Feedback

Participants provided feedback in response to questions on the post-session questionnaire (Appendix H). The answers here are used to inform us of possible improvements for the design of future studies.

User 1 confessed that he accidentally added a “P” value to his “L” sum before correcting himself. He commented that the “L” and “P” values were displayed too close together on the schedule. User 2 thought the exact memory limitation was missing. If the memory capacity was large enough, it may be possible to leave everything scheduled as is. It would have been nice for him to see a graph of memory. User 3 was not sure whether or not his greedy algorithm was the best method for solving the problem. He mentioned he could likely solve this discrete programming problem better with a program. However given 30 minutes, he believed his method was good. User 4 stated that he was unsure whether or not it was possible to schedule the downlink at a later time. He also mentioned that the memory limitation is not precise and he did not know if the memory was full or not at the first scheduled downlink. User 5 took a couple of minutes more than expected puzzling over whether or not it was possible to take pictures in the areas where the satellite was recharging power. In fact, he requested assistance part way through the task to ask this question. For our particular task, this would not have been possible. User 5 would have liked to be given a function of how the period of a picture depended on the cost of it.
For instance, is an activity 20 seconds long worth more than two consecutive activities 10 seconds long? In this experiment, they were worth the same value. The maximum memory storage would also have been helpful since User 5 could only guess at what that might be.

7. DISCUSSION

The purpose of this research was to study user annotation behavior in satellite scheduling. The results suggest several implications for the design of mixed-initiative schedulers. We believe that these implications not only apply to scheduling satellites, but also to scheduling in general. Again, we are careful not to infer too much about the user’s cognition and our analysis is based on mostly descriptive results.

This section is organized as follows. Implications for the design of mixed-initiative schedulers and the limitations of the experimental protocol are discussed first. Then comments are made about the markings and tools used. All this helps address our first and second research questions: (1) how do people use annotations to help them understand printed plans that are given to them and (2) do people use specific annotations because of the connotations of the annotation appearance or are arbitrary marks used? Comments are then made about the experimental task and suggestions for a future task are given. Then, some short comments on User 1 are made, followed by a discussion about some of the problems encountered in the study.

7.1 Implications for the Design of Mixed-Initiative Schedulers

This study shows that the use of marks helped in solving the problem. Although annotating paper is not the same as annotating machine schedules, I believe that people will tend to transfer their annotation practices from paper over to the computer. This was evident in the study by Obendorf (2003). As long as the task is complex and involves active engagement, people will need to use annotations to help them solve scheduling problems.
In terms of why certain markings were used in our study, many of the participants’ goals were found to be the same (see Table 1). In particular, everyone marked priority changes, deleted activities, wrote text, and either marked an interesting area or point on the schedule. Consequently, we believe that there is a need to provide annotations to support these four goals. At the very least, there should be a way to mark priorities. This could take the form of coloring an activity to indicate that it is frozen in or frozen out. Users want to keep high priority activities and delete less important ones in the next updated schedule. Another way is to provide a tool in which users can select from a choice of basic markings to add above or below an activity. These markings could resemble some of the ones used in our study such as vertical bars, checkmarks, arrows, and “+” signs. A possible distinction can be made between the freeze in and freeze out marks by only allowing freeze in marks to be added above or below an activity, while only allowing freeze out marks to be added over an activity. For instance, a big slash or cross over the activity would indicate the activity should be removed. These marks were seen on many of the schedules in our study. Text capability, which allows users to associate text to an activity on the schedule, a position in time, or an area that the user marked could also be valuable. These were all seen in our study. Finally, there needs to be support for marking interesting places, whether it is an activity, an attribute of an activity such as its label, a region between activities, a position in time, or anywhere else on the schedule. In our study, participants used vertical lines and arrows to mark positions in time and circled activities of interests.

7.2 Limitations of the Experimental Protocol

Because this was a paper study, participants could recommend changes to the schedule but they could not view an accurate representation of the results of these changes. They tried to predict power changes by drawing over the power graph and corrected their annotation mistakes by scribbling them out. These marks would not have occurred if this study were done on a computer. An algorithm would be able to compute power changes and mistakes would be
corrected with an "undo" command. If a program was used for the same task in this study, there would not only be fewer annotations used but the appearance of annotations would also be limited to what the system provided. Whether or not users will use annotations at all will also depend on the interaction technique. If annotations are hard or take too long to add, which may be the case for adding text, users may not bother to use them. This was found to be the case in our first pilot test with the scheduler. One of our users complained that adding annotations took too long so they were not used.

If the task was different, participants may have annotated differently. The task was simple enough for a 30-minute paper study but would be considered too easy for a computer study. If the task was harder and more time was given, there could have been more marks made. If there was conflict in the schedule where an activity had to be deleted first before a new request could be added, the annotations for doing this will look different from simply adding a new activity to an available time slot.

If participants had been asked to explain their answers to the interviewer at a much later time, additional marks may also have been made. Every participant declined to take a five-minute break after the task and User 3 stated that he wanted to precede right to the interview since “at least that way he would remember what he did.” If participants were told that they would be interviewed a day later, the annotations used may have changed. Clarification marks such as written notes may have been used to help participants remember what they had done if they needed to return to the schedule at a later time.

If participants were asked to justify their decisions to someone of authority that was going to judge their work, the use of annotations may also have been different. For instance, if their performance were going to be graded, perhaps the marks they make would appear tidier and contain more explanatory notes. In the real world, it is conceivable that machine schedules sometimes do get saved for further examination at a later time or that schedule analysis is
performed in collaboration with someone else. Annotation practices may likely be quite different for these particular instances.

The generalizability of our results only takes into account how the majority of people behaved during our task and only considers the actions that cannot be performed using a computer. As mentioned in the above section, we believe support is needed for marking priority changes, deleted activities, interesting regions, and adding text on the schedule. An algorithm can compute the effects on power, correct for annotations, draw in new requests, and test solutions.

7.3 Markings and Tools Used

The different markings used to mark the activities that have doubled in value seem to support the idea that annotations were personal and somewhat arbitrary. Personal annotations are only meant for the original annotator to understand. All the marks are somewhat arbitrary except for User 1’s markings. His 2 inside the square below the activity symbolized that the activity had twice the value of a basic image. User 2’s markings with an inverted arrow symbolically seems to contradict the fact those activities have increased in value. Interestingly, in one of the earlier pilot studies, a participant used an upward arrow to mark his priorities and mentioned it was used to indicate that those activities have increased in value. Another example that appeared to contradict logic was when User 3 used a red “+” sign to indicate the deletion of an activity. The “+” sign is normally associated with addition and one would have likely interpreted this to mean that the activity was being added or kept in the schedule. It would have been more understandable if a “-” sign was used instead. Symbolic or meaningful annotations can further help boost recall of learned material but the results here appear to show that arbitrary markings can also be effective. These findings also contradict our hypothesis that in most cases, people will use specific annotations because of their connotations. However, as the problem grows in size and complexity, arbitrary markings alone may not be enough. Also, if the user plans to return to the task at a much later time, arbitrary markings may not be very useful.
The labels placed inside the new request box seemed to portray meaning. User 1 used “new”, which one can interpret as representing the new request (Figure 5). User 2 used “R” perhaps to indicate it was a request or region since region R was specified in the problem (Figure 6). User 3 used “F” which is the first letter of the word “fire” or “forest fire” (Figure 7). The new request was to observe a forest fire (see Appendix E).

Some of the tools used appeared to be arbitrary. Two people used a blue pen and three people used a black pen. When asked if there were any particular reason for choosing the particular pen, everyone but User 3 answered that there was no reason and that it was the first pen they grabbed. User 3 mentioned that he normally uses black pens or pencils for working things out. Although most people responded “no reason” to this question, they may have had reasons for not using another tool such as a red pen. Many people use pencils or blue pens to brainstorm about problems since red pens are associated with grading.

7.4 Comments about the Task

The problems specified in the task were formulated with mixed-initiative planning in mind. Because the task was only 30 minutes long, we could only test a small problem. All participants spent a considerable amount of time just reading and trying to understand the problem. The first draft of the task presented detailed descriptions of several new requests that were in conflict with each other on the schedule. Participants were required to use their judgment and choose which request to schedule and which to discard. This type of human involvement is typical of mixed-initiative interaction and cannot be computed by an algorithm. Unfortunately, this task had to be simplified since it took nearly one hour to complete in the pilot test. The task used in this study posed an easier problem involving only prioritizing activities based on value alone. In this case, a computer could be used to perform the prioritization.

The task in the user study still presented several changes to the schedule that could not have been predicted by any machine. Unless a computer is equipped to handle such circumstances and
the program is constantly maintained, human assistance is essential. The goal of the task was to have participants make tradeoffs on different priorities such as total value, power level, and data capacity. Humans are superior at making trade-offs between goals, while computers are better at performing quantitative analysis of proposed courses of action (Hearst, 1999). For example, the participant chooses which activities to freeze in or freeze out and then the computer performs the commands and reflects the changes. In our paper study, participants were marking their priorities and identifying their deletions and these would translate into freeze in and freeze out commands.

A task which forces the user to think about what may happen if several activities were added throughout the schedule would be a good task. When an activity is added, power is drained and that could lead to a lower priority activity being deleted. Only one addition was required in our user study and it did not require significant effort to solve. If many additions were required, it would be much more difficult. This scenario may be suitable for a user working with a computer program but is certainly unfeasible for a 30-minute paper study. If users could see the results of their action computed by an algorithm, they could then modify the results and improve their solutions.

7.5 Comments on User 1

Some of the notes that User 1 made appeared to be answers to the problem in the task. His notes clearly stated what could be done and why it was possible to do so under the constraints (Figures 10 and 11). The notes may have been taken as a way to share his answers afterwards in the interview, could simply have been made to help him along with the problem, or both. When asked whether marks were made as a way to share his answers in the interview, he claimed that they were not. It is possible that he may have only considered the markings on the schedule rather than on the sheets. However when asked about the markings he made, some of his answers did refer to markings on the sheets.
User 1’s response to some of the questions here gives us a reason to believe that the participants may not have accurately answered some of the questions. In this particular case, perhaps two separate questions should have been asked instead: (1) were any of the marks on the schedule made as a way to share your answers and (2) were any of the marks on the source documents made as a way to share your answers?

7.6 Problems and Issues

Even at full quality, the video recordings did not reveal details of the markings on the schedule. The camcorder was positioned so that we could capture the participant’s interaction with the schedule. If the participant leaned forward while making marks, their facial expressions would also have been captured. Their reaction was not as important to us, which is why the camera was focused on the schedule and surrounding table (Figure 4). At times, a participant’s arm or hand would block the view of what had been noted down. This was especially the case when the participant was right-handed. The camera was placed to the right of the table out of convenience due to the orientation of the room. For future paper studies of this sort, multiple cameras, including one mounted to the ceiling, could be used to capture different angles of interaction. For this study however, I did not believe that resolution was very important. We were more concerned with the types of markings that were made than with the order that marks were drawn in, exactly how they were drawn in with a pen, or exactly where the participant was gesturing. The videos showed roughly which part of the schedule was being interacted with, what tool was used, whether the participant was reading or marking, and the audio was clear. For our purposes, details from the videos were not required and all the information gathered was sufficient.

The explanation sheet on satellite scheduling did not explicitly state that activities on the schedule could not be moved within the schedule. User 3 and 4 both thought that they could schedule the downlink once it returned. This would have been a valid assumption with other
types of scheduling but is not the case with scheduling satellites since satellites are in orbit around the Earth.

The last question in the interview about how the use of marks helped participants make a decision appeared to be too broad. Either short answers were given such as the marks made the priorities clear, or in the case of User 2, no answer was offered. All the users agreed that the marks helped them solve the problem. I believe that the question may have been a hard one to answer without much thought. Asking more specific questions about how marks helped and allowing them to answer yes or no may prove to be more effective. For instance, “Did you use marks to interpret and comment upon the schedule?”

8. DIRECTIONS FOR FUTURE WORK

One of the goals of this study was to make recommendations for the design of the scheduler implemented by Brown and Kirkpatrick (2005). A follow up study should be conducted using the scheduler but before that can happen, many changes have to occur to the software and design protocol. The interface of the scheduler needs to be modified to resemble the paper schedule. In particular, the length value of each activity, the power scale, and horizontal grid lines on the power graph should be displayed. All this information was useful and providing them allowed participants to focus on the task of making decisions rather than spend a lot of time making computations. I also believe that the second panel display from the top on the original schedule (Figure 1) is not required and should be removed. In fact, in the first pilot study with the scheduler, one of the participants had mistaken the power graph as corresponding to the activities in the second panel rather than the third panel. In addition to correcting the current bugs in the scheduler, two new functions should also be added. It should be possible for a new activity to be added to the schedule and for annotations to be associated with times. Currently, annotations can only be associated to activities. Also, freeze in commands are valid for all future updates whereas
freeze out commands are valid only for one update cycle in the current version. It is still unclear why this was the case but both commands could be made valid for all future updates to minimize confusion.

In terms of the design of the next user study, a few suggestions can be made on the type of tasks and questions given out. A task involving the additions of three or four new requests throughout the scheduler could be used. The requests could be described in detail and should be chosen to conflict with at least one other activity. Participants should also be told that all initial activities on the schedule can only be deleted and not moved. This can also be stated on the satellite scheduling explanation sheet (Appendix D). The participants would have to first prioritize the activities before making changes to the schedule. Commands for additions, deletions, and activities to keep could be issued to the scheduler and it would compute and show the resulting solution. Because this sort of interaction is possible with a program, the participant can contribute to the iterative and evolving process of problem solving. If this task cannot be completed in one session time, then two sessions should be used. This task clearly demonstrates the interaction that is typical of mixed-initiative scheduling.

Because participants appeared to have some difficulty answering how the marks they used helped them in solving the problem, more specific questions should be asked with the option of answering yes or no. For instance, “Did your marks help you understand the data you gathered?” Did your marks help you remember what your priorities were? Did you make marks as a way to work out problems?

As a final suggestion, we could request for participants to think-aloud during the task. This may allow us to better understand why they made certain choices. The potential problem with this is that participants may feel self-conscious since thinking aloud requires people to construct rational stories about their behavior. In the case of solving scheduling problems, a participant’s
behavior may be exploratory while he or she is trying to comprehend the task; there may not always be a strict rational reason for every action that is taken.

9. CONCLUSIONS

Interactive systems for solving planning and scheduling problems are becoming pervasive in many application areas such as space missions, rescue, and vehicle routing (Cortellessa et al., 2004). In most practical domains, most systems do not support the iterative, evolving process of problem understanding, requirements specification, conflict resolution, and solution improvement (Smith et al., 1996). To be effective, the designs of advanced interactive interfaces need to integrate the human operator in the problem solving process. For tasks that require human cognitive abilities at a higher level of abstraction and decision, annotation tools could provide an aid, which preserves the “traditional” real world practice.

In this study, we investigated how people used annotations to solve problems when presented with printed satellite schedules generated by a computer. Participants were observed on video and the answers from the interviews were correlated with the marks on the schedules. The following summarizes the answer to our first research question. While making trade-offs on different priorities, all five participants in the user study benefited from the use of marks. Our participants used annotations to help them understand plans by using marks on the schedule to emphasize high priority activities, to mark “interesting” regions, to indicate deletions, to draw and label a new request, to compute the effects on power, to associate notes to places of interest, and to test solutions using pieces of paper. Some source documents were also annotated to highlight text, to write out answers, and to perform calculations. Every participant approached the problem in a similar fashion; marks were made as reading took place and the task, reference, and schedule sheets were constantly being referred to throughout the task. The black or blue pen and calculator were the most common tools used. User 3 used more tools than anyone else and interacted the
most with the schedule. User 3 was the only one that claimed some of his markings were for sharing purposes. One person thought the task was easy; another thought it was hard, and three people thought it was reasonable.

Our second research question was whether people used specific annotations because of the connotations of the annotation appearance and how meanings were assigned to the marks that were used. Besides User 1, it appeared as though participants did not use specific annotations because of the connotations of the appearance. The marks used to identify high priority activities were arbitrary. User 1 marked activities that have doubled in value with a 2 inside a square. User 3 used a red “+” sign to indicate a deletion whereas others used a slash or X through the activity. Everyone did assign meanings to their marks but because they were mostly personal, many of them could not be interpreted.

To address our third and final question on making recommendations for the design of the scheduler implemented by Brown and Kirkpatrick (2005), we highlight the changes that need to be made. The next step is to conduct a follow up study using the scheduler. Before that can happen, bugs need to be fixed, changes need to be made to the interface, and some new functionality needs to be implemented. If possible, a task typical of mixed-initiative interaction should be used. The questions and forms used in this study can be used but modifications should be made to include specific questions about how marks were used rather than have participants explain in their own words how their marks helped them make a decision.

Although annotating paper is not the same as annotating machine schedules, I believe that people will tend to transfer their annotation practices from paper over to the computer. Results from this study show that support is needed for marking priority changes, deleted activities, interesting regions, and adding text on the schedule. I hope the results will provide future designers with some insight on how to build better interactive satellite scheduling systems or any scheduling system, for that matter, with annotation capabilities.
REFERENCE LIST


APPENDICES

Appendix A. Agenda

Satellite Scheduling – Pilot Study Summer 2007

Welcome
Consent form
- Procedure: “You will be given a schedule and asked to read and solve a problem. You may then take a break, if you wish before I interview you about your answers.
- Possible risks: none
- Benefits of research: to understand how problems are solved in satellite scheduling in order to build better systems.

Background question sheet

Satellite Scheduling description
Explain schedule and reference sheet:
- “Read the entire problem first. We are not concerned about accuracy or correct answers. Rough answers will be acceptable. Feel free to write on any of the sheets I gave you. Do the best you can with the information that is given. I will be close by if you really need help.”

*Provide scrap paper, color pens, color highlighters, post-its, ruler, calculator, & timer/watch

Start recording for video. Reset timer and start countdown.

Task – Problem | 30 mins.

Stop video.

Break | 5 mins.

Start recording for video.

Interview:
- Justification for Recommendations
- Post-session questionnaire
- “Check schedule for any illegible marks”

*Take rough notes for answers
Stop recording for video. THE END.

Escort participant out and provide payment.
Bring in next participant.
Appendix B. Consent Form

School of Computing Science, Simon Fraser University

Title: Understanding Satellite Scheduling
Investigator: Susan Villecroze

The University and the members of this research team have designed this study to protect at all times the interests, comfort, health, safety, and psychological well-being of participants. This research is being conducted under permission of the Simon Fraser Research Ethics Board. This study has been designed and is being conducted according to the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association.

Any information that is obtained during this study will be kept confidential to the full extent permitted by the law. You will not be required to write your name or any other identifying information on research materials. This consent form will be the only document containing your name. It will be kept separately from all research materials and there will be no means of connecting this from with the videos and notes from your session. All research materials, including videotapes, written records of the study, and consent forms will be maintained in a secure location.

Your signature on this document indicates that:

- You have received and read a document that describes the procedures, possible risks, and benefits of this research project and have been given adequate opportunity to consider the information in the document.
- You voluntarily agree to participate in the project.
- You may withdraw your participation at any time with no penalty.
- You may receive a copy of this form.
- You are aware that the study will be videotaped, that your name will not be recorded in the video or any related written documents, and that the video will be maintained in a secure location.
- You have the choice of having the video of the study in which you participate being destroyed after a period of 2 years, or allowing the video to be maintained by Arthur Kirkpatrick, who is supervising this study.

Please initial one option:

- [ ] Video must be destroyed after 2-year period
- [ ] Video may be maintained by Dr. Arthur Kirkpatrick

- You may request copies of the result of this study, upon completion, by contacting Dr. Arthur Kirkpatrick (ted@cs.sfu.ca), or Susan Villecroze (svillecr@sfu.ca).
- You may register any complaint with the Director of the Office of Research, Dr. H. Weinberg (hweinber@sfu.ca), the Director of the School of Computing Science, Dr. J. Delgrande (jim@cs.sfu.ca), or the researcher Susan Villecroze (svillecr@sfu.ca).

Name (print) ____________________________________________  Last ____________________________  First ____________________________

Signature ____________________________________________

Date ____________________________  (DD/MM/YYYY)
Appendix C. Background Information Form

User #: 

Age: 
Gender: M / F

1. Please indicate the courses you have taken.

☐ Linear algebra
☐ Integral calculus
☐ Differential equations
☐ Discrete mathematics
☐ Statistics
☐ Real analysis
☐ Set theory
☐ Computational complexity
☐ Algorithms and analysis
☐ Other: ____________________________

2. How would you rank your mathematical abilities?

☐ Very Strong
☐ Strong
☐ Good, but not strong
☐ Weak

3. When reading a document, plan, or schedule, under what circumstances (if any) do you mark up the document?

4. Have you ever worked with or studied the scheduling of satellites?

☐ Yes
☐ No

5. If you answered yes to #4, please give details.

6. What is your energy level like today?

☐ Energetic
☐ Normal
☐ Fatigue (Tired)
Appendix D. Satellite Scheduling Explanation

Earth Observing Satellites earn money by taking images of specified locations on the Earth surface, in response to observation requests. Each requested image is assigned an image request ID and can be scheduled at several points in the day.

Images come in two types: high-priority and low-priority. While both types of images earn money, high-priority images are worth more money than low-priority images, if all other things are equal. Images that require more time to schedule (i.e. longer images) are worth more than those that require less time (i.e. shorter images).

Satellite scheduling involves the creation of a “plan,” which essentially tells the satellite what to do at any given moment. Not only must satellites take images, time is scheduled for recharging the batteries, during which no other actions may be taken. Time may also be required to transfer data down to a ground station on Earth during a downlink connection. It is best to downlink when the satellite’s memory storage is full, but this cannot always be the case.

Because the number of image requests generated each day will exceed what can actually be scheduled and can number in the hundreds, individuals who do satellite scheduling require the aid of a software planner. Earth-observing satellites also need to respond to unexpected changes in weather, natural disasters, and changes in satellite or ground station capability. Because of this, it is necessary for users to interact with the system at different levels of automation. In this iterative process, an initial plan is presented, problems are recognized, constraints and requirements are adjusted, the solution is revisited and so on.

In this study, you will evaluate a printed schedule, which has been pre-computed for you. The value of each schedule is displayed at the top of the page. The length of time that is required to observe the target and its associated priority determines its value.

You will have 30 minutes to complete two problems. When the 30 minutes is up, the task will end and you will be asked to explain your decisions in an interview.
Appendix E. User Task

Note: Your sample schedule has been pre-computed and only contains basic images (low-priority tasks). The following constraints and considerations should be taken into account when making recommendations: power, memory limitations, task priority, and total plan value.

Problem Scenario

There are two downlink opportunities from two different locations in the satellite’s orbit. These tasks are indicated by patterned-looking rectangles in the schedule. They are high-priority tasks (29, 30) and (54, 55).

At time 11:35AM, an alert message arrives stating that there is a problem with the ground station, which you have a scheduled downlink at task 29 and 30. It cannot receive any downloads for two hours (that is how long it will take to fix the problem). You will definitely miss the downlink opportunity. This means you only have one downlink opportunity. Your schedule currently assumes the satellite’s memory storage is empty after task 30. Shortly after, you are notified that the following tasks are now worth twice as much as before: 10, 15, 24, 33, 37, and 43-46. On top of that, you just received a new request. It appears that the new request is worth 3 times that of a basic image.

New request: A forest fire has just started in region R and needs to be observed. Region R will be approached by the satellite and can be photographed between 1:28PM and 1:30PM. The observation should last at least 20 seconds, but no more than 30 seconds.

Your task is to recommend modifications to the schedule while still operating within specified constraints. See if the new requests can be added and if so, what needs to be done, in terms of deletions and additions to the schedule? Also, assume at the start of the schedule the memory storage is empty. Note: It is not unusual for requests to be left unfilled.

Summary of issues to take care of:

1. You will have to reconsider the schedule now that memory will NOT be empty after task 30
2. Some tasks are now worth twice as much as before
3. You received a new request, worth 3 times as much as a basic image
Appendix F. Reference Sheet

*Power constraint:* You must maintain the battery reserves above a minimal level. Remaining battery power is displayed in the bottom panel of the display, on a scale from 0 to 100%. The dashed line partway up indicates the minimal reserves. Your goal is to maintain the battery reserve above the minimal level so that the battery power does not go below that level at any point in the plan.

Max power (100%) = 180 units of power
Min power at dashed line = 30 units of power

*Total plan value:* The idea is to schedule as many tasks as possible so as to maximize the total value of the plan sent to the satellite. Remember that higher priority images and longer images are worth more.

*Memory:* Hint - think of how the number and length of the tasks could be associated with the memory limitation.

You will likely have to delete basic images from the schedule you selected in order to fit in other images. Note, when you delete a basic image, not only does it free up power, but when no other image is scheduled in its place, it also enters the satellite into a recharge state and gains more power. You can assume that the satellite will not charge past the maximum level, even if it is in a gain state. Because we are always either in a power drain or power gain state, if you must choose between an isolated task and two consecutive tasks to delete, choose the isolated task.

*Note:* adding a high-priority image will draw power at the same rate as a basic image.

Power drain rate = 1.5 * (length of task in seconds)
Power gain rate = 1.0 * (length where there is no activity in seconds)

*Opt for safety:* when in doubt of the power level, leave images out or delete more images.
Appendix G. Interview Questions

User #:

Recommended Changes: Interview Questions for Participants

Justification of answers

1. If you were able to add the new request, describe how it was done. What were your reasons for deletions for the new request?

2. Did you prioritize the tasks on the schedule? If so, describe how you did it.

3. Describe your strategy for solving the downlink problem.

Use of “marks” in scheduling

1. Did you use any marks to note places you might want to return to in the schedule (like a bookmark)?

2. If you assigned meaning to your marks, how were they assigned?

3. Did the meaning you assigned to a mark change over time?

4. Did you make marks as a way to share your answers afterwards?

5. Was there any reason you picked the tools that you used?

6. Explain how your use of marks helped you make your decision on which tasks to keep or remove?
Appendix H. Post-Session Questionnaire

User #:

1. Did the use of marks help you solve the problem?
   - Yes
   - No

2. Was there anything that you needed in the experiment but was not provided for you?

3. How did you find the problem given to you in the experiment?
   - Too easy
   - Easy
   - Just right
   - Hard
   - Too hard

4. If your answer in #3 was hard or too hard, what parts of the problem did you find hard?

General Comments or Suggestions: