IPRO 311 Midterm Report

IPRO 311

1 Planning

1.1 Objectives

The Spring 07 semester of IPRO 311 will focus on the task of implementing a system for capturing and comparing user habits in order to create a misuse detection system that determines if a user is using a computer for nefarious purposes. But what is computer misuse? Generally, when people think about computer crime, they think of the typical 14-20 year old hacker, sitting in his parents cold, dark basement breaking into a CIA mainframe. And while this is a serious problem, it isn't the most common computer crime. Exceedingly more common is the computer misuser, who can fall into two categories. The first is your typical "slacker" who plays flash games all day instead of doing work. The other, more dangerous computer misuser, is the one who abuses their privileges, accesses generally restricted information, and uses it for nefarious purposes. For example, a higher-up exec in a weapons facility is allowed to look at the research for the newest weapon system. Him taking that information and selling it to competitors would be an example of misuse.

Our objectives have not been substantially altered, as our goals have not changed. The reason behind this is the true intent has not been altered.

1.2 Task Definition and Durations

NOTE: please see attached Microsoft Project document for the updated WBS. Included in the document are updated summary tasks and hourly durations. For individual tasks, please see §2.5.2

2 Organizing

2.1 Accountability

Team Leader – Yacin Nadji
Coding Team Leader – Jason Soo
Design Team Leader – Justin Choriki
Query Processing Team Leader – Jong Min Lim

2.2 Coding Team

The coding sub-team is responsible for two major things. First, they need to develop the automated query logging system to record the following:

Items to record at the Operating System level

* Snapshot of files+binary diff of daily changes
* All keystrokes

  Saved to a file(keylog.txt) with a timestamp every minute

  ```
  Example:
  200701051231:this is test text, this is test text.....
  200701051232:
  200701051233:some more test text after 2 minutes
  ```

* All newly saved files(full listing and actual file)

  Files will be saved in folder saved_files/ with format filename.[usb|floppy|cdr|hd].version, the files will retain the original permissions and ownership as original file

  ```
  Example:
  ls -lh
  -rw-r--r--   1 jwilberd jwilberd   18 Sep  8  2006 testfile.hd.1
  -rw-r--r--   1 jwilberd jwilberd   18 Sep  8  2006 testfile.hd.2
  -rw-r--r--   1 jwilberd jwilberd   18 Sep  8  2006 testfile.hd.3
  -rwxr--r--   1 jwilberd jwilberd   18 Sep  8  2006 testfile2.usb.1
  ```

* All newly opened files(full listing and actual file)
  * Format will be the same as saved files only folder will be called opened_files/
* All newly created files(full listing and actual file)
  * Format will be the same as saved files only folder will be called created_files/
* Running processes
  * Stored in process_list.txt, with timestamps
  * Initial process list
  * New processes
  * Ended processes
* Memory usage
  * Stored in memory_usage.txt, with timestamps
  * 1 sec intervals

* CPU usage
    * Stored in cpu_usage.txt, with timestamps
    * 1 sec intervals
* Disk usage
    * Stored in disk_usage.txt, with timestamps
    * 1 sec intervals
* Mouse movments

  Record mouse position x,y every second to mouse.txt with timestamp

  ```
  Example:
  20070105123101:100,201
  20070105123102:567,231
  20070105123103:120,350
  ```

* Raw network data
    * Save to network.txt using pcap's internal format

Items to record at the Information Retrieval level

* The queries
    * Order of queries
    * Time between queries
    * End of session indicator
* Top retrieved documents(20-100)
    * HTML
    * URL
    * Document Ranking
* Click through data
    * Length each link was visited
    * Amount of text was read(scrolled)
    * Order documents were clicked
* External links followed(capture same data as initial link)
* User assisted input
    * User's subjective ranking of documents
    * Record the nature of each task
    * Record purpose of each query

* Each system event will be recorded in a database
    * date
    * time
    * user
    * event type
    * time since last system event
    * time since the last system event of this type
    * query id
* Each system level event can be then mapped back to the query whose results the user was viewing when the event occurred based on the user, date and time the query was issued

After these necessary tools have been implemented, the sub-team will need to code a series of scripts to present the information from several different angles, most notably one allowing the Query Processing Team to make use of the data to run a series of ROC curves to analyze progress. It's also necessary to present the information logged in such a way it will be presentable to anyone at any education level.

Members:
Jason Soo
Yacin Nadji
William Alton
Matt Holmes

2.3 Design Team

The Design Team's responsibility is to design and implement the website, brochure, poster, and final presentation. The individuals were selected based on their technical merit in the design field, and ability to succinctly convey information. The sub-team tasks are not as time consuming as the other sub-team's, so they will help the other sub-teams as necessary throughout the course of the project.

Members:
Justin Choriki
Mark Malanowski
Daniel Hyc
Hee Yeol Jeong

2.4 Query Processing Team

The Query Processing Team is essential to taking all the collected data, throughout the semester, and accomplishing 3 main tasks:

1. Make sure the logging system is functional

2. Analyze the data, searching for patterns between proper use and misuse
3. Present the findings weekly to the team

Members:
Jong Min Lim
Young Cho
Peter Niedzinski
Gerardo Sanchez
Jong Mu Song

2.5 Role and Resource Allocation

2.5.1 Budget

No budget is necessary for the completion of this project. All code not directly developed by the teams have been complimented by various Free Software solutions.

2.5.2 Roles Allocated to Individual Members

Name	Major	Skills & Strengths	Role & Tasks
Yacin Nadji	Computer Science	Programming, Operating Systems, Public Speaking, Management	IPRO Team Leader, Responsible for implementing OS level logging and maintaining the Knoppix bootable CD
William Alton	Computer Science	Programming, Data Mining	Develop IR Level logging and maintain the code-base
Young Cho	Applied Mathematics	Statistical Analysis, Databases	Providing up to date analysis of current query records
Justin Choriki	Mechanical Engineering	Art, Design, Webdesign, Layout	Design Sub-Team Leader, oversees completion of the web site, brochure, poster and final powerpoint presentation
Matt Holmes	Computer Information Systems	Python scripting, text parsing	Parsing log files into presentable format for the Query Processing sub-team
Hee Yeol Jeong	Electrical Engineering	Mathematical and statistical analysis, design, physics	Handling the presentation of the statistical information for the design team. Acts as a liaison between the Design and Query Processing team
Jong Min Lim	Electrical Engineering	Leadership, statistical analysis, circuit design, applied mathematics	Leader for Query Processing sub-team, oversee query processing and participate in statistical analysis
Mark Malanowski	Computer Science	PHP/MySQL, dynamic web programming and design, scripting languages	Back-end web design (PHP/MySQL) and scripting support for Coding team
Peter Niedzinski	Biology	Public speaking, writing, applied mathematics	Aid in statistical processing and present the information to the team on a regular basis. Provide textual material for IPRO Day. Taking the minutes.
Gerardo Sanchez	Civil Engineering	Applied mathematics and statistical analysis	Aid in statistical analysis of the results
Jason Soo	Computer Science	Scripting, data processing, information retrieval	Coding sub-team leader, oversee and develop scripting for data collection
Daniel Hyc	Computer Science	Web design, scripting	Head web developer

3 Controlling

3.1 Results to Date vs. Original Plan

Since the original Project Plan, several changes have been made to the project management aspect of the IPRO, but the course and successful completion of our goals has remained well on course. We have completed the query logging system, and have also finished the testing phase for the query logging system. At this point, the entire team, in addition to their regular tasks, will also run 20 queries each week in order to create our data set for the ongoing analysis, and data set generation.

As can be seen clearly by the Gantt chart, there has been a slight deviation in the completion of the website and the poster. Several of the core web/poster design group also have extensive coding experience, and in order to get the logging system up and running on time, it was necessary for them to help the Coding sub-team finish up the logging system on time. Considering how much time we have to finish up the website and poster, the team feels it correctly re-adjusted our time to aptly handle the problem at hand.

There was mild discussion as to when we should begin the query processing phase, if we started too early, we might get a poor view on the current state of the data and subconsciously change our query/misuse habits, start too late, and we might miss potential problems in the query logging system. Neither of which seem like too good of a consequence. We decided to alter the plan a bit, and have the analysis begin immediately. It seemed as though allowing to see the results would give us a better understanding of what we were doing, and much like individuals in the workplace, we would change our habits. Originally, we believed this behavior to be a negative one, but in retrospect, it will provide us with a more accurate representation of a data set obtained from a legitimate industry workplace. In order to keep our goals a reality, it's necessary to keep strict to the schedule.

3.2 Monitoring of Project Status

The current obstacles are essentially generating enough data, and organizing it in such a way to run the necessary Data Mining algorithms to recognize abnormalities when misuse has occurred. Naturally, we may run into server trouble or run into faulty code, which is why the coding team's main responsibility is to keep the code-base and the servers up to date and free of bugs. Continuing IPROs will use the software suite developed during this IPRO, so it is requisite to continue testing the software.

Future obstacles are more difficult to guess, considering it's nigh impossible to predict everything that will happen in a project life-cycle. Potential problems we foresee involve queries not being logged properly, and after gathering all the necessary information, the information isn't sufficient enough to recognize patterns. The solutions to both of these are simply keeping up with our current plan, making sure everything is working each step of the way.

4 Contacts

The IPRO 311 team consists of the following members:

Last Name	First Name	E-mail	Phone	AIM
Frieder	Ophir	frieder@iit.edu	[redacted]	none
Goharian	Nazli	goharian@iit.edu	[redacted]	none
Grossman	David	grossman@iit.edu	[redacted]	davhippo
Nadji	Yacin	ynadji@iit.edu	[redacted]	G7 Onizuka
Platt	Alana	platala@iit.edu	none	dreamingdaphne9
Wilberding	Jordan	wilbjor@iit.edu	none	diginux0

5 Document Submission

When submitting documents, we have three locations that the documents must be submitted to, with the appropriate person defined for moderating it.

Submit Location	Website Address	Person Responsible	E-mail
Website Wiki	zchor.iit.edu	Yacin Nadji	ynadji@iit.edu
iKNOW	iknow.iit.edu	Mark Malanowski	malamar@iit.edu
IGROUPS	igroups.iit.edu	Mark Malanowski	malamar@iit.edu

6 Meetings

There will be two required meetings per week, one main meeting for our general work, and one mid-week checkup, to keep the team on track.

* Mondays 17:00-19:40 E1 027

* Thursdays 21:30-22:00 AIM, A brief mid-week discussion

7 Weekly Submissions

Each team member is required to submit a two paragraph document each week by 11:59PM Sunday, that details what they worked on (number of hours for each), problems they encountered, and their plans for the next week. The document should be submitted to blackboard.