Show Me How You See: Lessons from Studying Computer Forensics Experts for Visualization

T.J. Jankun-Kelly[1], Josh Franck[2], David Wilson[1], Jeffery Carver[3], David Dampier[1], and J. Edward Swan II[1]

[1] Department of Computer Science and Engineering, Mississippi State University {tjk,dwj12,dampier,swan}@cse.msstate.edu
[2] Department of Psychology, Mississippi State University jaf210@msstate.edu
[3] Department of Computer Science, University of Alabama carver@cs.ua.edu

Abstract. As part of a Analyze-Visualize-Validate cycle, we have initiated a domain analysis of email computer forensics to determine where visualization may be beneficial. To this end, we worked with police officers and other forensics professionals. However, the process of designing and executing such a study with real-world experts has been a non-trivial task. This paper presents our efforts in this area and the lessons learned as guidance to other practitioners.

1 Introduction

While violent crimes such as armed robbery and murder are decreasing in the U.S., computer crime is growing world-wide [1–3]. The growth of the Internet has contributed to an increase in cyber crimes such as child pornography, gambling, money laundering, financial scams, extortion, and sabotage [3–5]. Besides their using a computer in the commission of a crime, computer criminals share another similarity: The chances of their being caught and successfully prosecuted are relatively small [1]. In one example, a sheriff's department investigator working exclusively on computer crimes full-time for five years made only five arrests, none of which led to convictions [6]. Thus, the need for tools to alleviate the workload of computer forensics practitioners is clear.

Several commercial and open-source tools exists to assist forensic officers. While these systems automate some tasks and facilitate others, we believe there is room for visualization to be of assistance. Some initial efforts have been made in this area [7]. However, the work process of computer forensic officers and the interaction with their tools has not be thoroughly studied; we do not want to use visualization to solve problems which are irrelevant or done better by existing tools. We have therefore initiated a domain analysis of law enforcement computer forensic personnel to discover where forensic visualization may be fruitful. In this paper, we discuss the iterative process we went through when working with these experts, some preliminary results, and the lessons learned from this effort. We hope this contribution will serve as a case study for future similar efforts in security visualization in a similar vein as other case studies in visualization and software engineering [8, 9].

2 Case Study: Webmail Forensics Domain Analysis

Computer crime takes on many forms; a study of forensic analysis of all such crimes is beyond our scope. Thus, we focused on one aspect of computer forensics: Webmail and Internet history analysis in fraud cases. The ability to generate numerous web email accounts and the difficulty of putting together a coherent timeline of webmail usage motivated our investigation. In addition, finding corroborating evidence from other files on a hard drive (e.g., financial records) was of interest. Existing forensic tools such as the Forensics Toolkit [10] and the Autopsy Forensic Browser [11, 12] provide interfaces to search files and unallocated sectors for relevant information, but do not necessarily provide guidance on what to search or how best to find evidence. The effectiveness of finding evidence with these tools is thus highly practitioner dependent. Therefore, we wanted to study the workflow of such users at different levels of expertise to discover how visualization could benefit this task.

2.1 Study Protocol and Design

Our study's design evolved over several iterations as we interacted with our experts, though the eventual visualization goal was kept fixed over each version. We initially considered a full verbal think-aloud protocol design [13, 14]. Such analysis is one in which the recorded verbalized thoughts of subjects are broken into the smallest information-bearing segments and coded categorically (e.g., Search, Navigate). These coded segments can then be examined for frequency of individual coding types, for recurring segments of coding types, and for tracking the evolution of the subjects' cognitions over the course of the task. Most importantly, by connecting the time stamp of coded segments to the specific subtasks being performed during that period, these same connections can be drawn on a subtask-by-subtask basis. Unfortunately, this process not only requires a great deal of time and effort to perform, but generally produces an almost toxic level of data. Also, it is the most complex and data-rich analysis method available to examine this particular task. As the exact relationship between the data gathered on the subject-side and the changes made on the visualization-side has yet to be established, this method was deemed excessive for an initial study.

For our expert trails, we settled on a more open-ended contextual task analysis utilizing a greater range of data souces. During a contextual task analysis, the goal is to observe users within their normal work context, as opposed to an artificial setting such as a laboratory [15–17]. For this purpose, a laptop with the forensics software (Autopsy), test cases, and observational software (screen/mouse capture and a web camera) was provided to subjects (Figure 1 left.). Subjects were instructed to "do what they do normally," and were provided (unobtrusively) with a notepad and pen with which to take notes, should they desire. There was also an embedded note-taking system within the analysis suite used. These two note-taking sources were intended as the primary source of data, and were to be analyzed in a fashion similar to the verbal protocols described above, though at a higher (and less cognitively complete) level. The

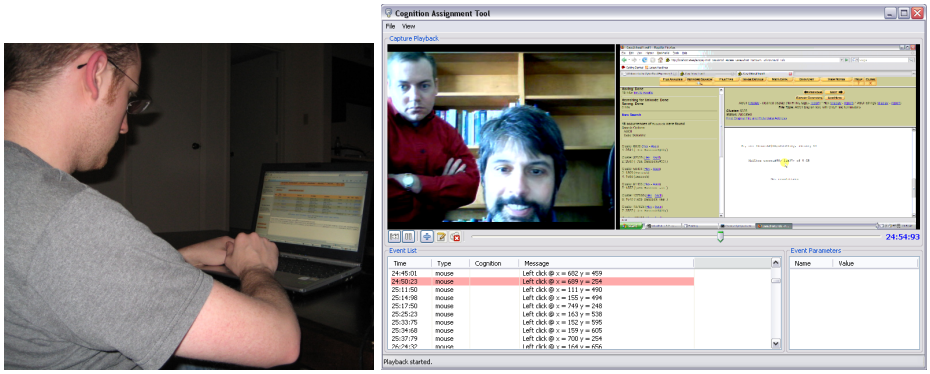

Fig. 1. Experimental setup for our task analysis. *Left*: Experimental rig. A version of Autopsy [11] modified for capturing user events was the forensic software used. *Right*: Analysis tool we developed to process expert trials. Webcam input is in the upper-left, screen capture video in the upper right, mouse events on the lower left, and observer notes and analysis on the lower right.

screen/mouse capture and audio/video data provided by the laptop were used to complement these notes. The screen/mouse capture data was intended to be examined for instances where activity indicated confusion, or where recurring patterns of input required an excessive amount of movement in terms of screen distance. The video and audio data were intended to be examined informally to inform the design of our subsequent visualizations; we use a custom developed analysis framework that syncs the video, screen capture, and mouse event display for this purpose (Figure 1 right).

While we wished to observe experts analyzing real cases, the legal and time concerns involved, especially for ongoing criminal investigations, forced us to create test datasets. Based upon our discussion with our in-house forensics instructor, we created two webmail fraud cases that mimicked attributes of similar cases. We interjected fraud related emails into email streams collected from active mailing lists traffic to which our test accounts were subscribed, and intermixed web-page views related to the fraud with visits to high traffic sites such as Google, Yahoo!, and Wikipedia. The rest of the hard drive images contained a standard Windows XP installation. Both our test datasets were used in the task analysis as described next.

2.2 Study Execution

Once the task analysis study was designed, we set about soliciting subjects. Initially, officers taking courses at the Mississippi State Forensics Training Center were considered; however, most did not possess sufficient enough knowledge to provide any meaningful "expert" data. Consequently, we solicited known forensics officers throughout the state of Mississippi and neighboring areas through

email and phone calls. This recruitment process was labor intensive, requiring on average 30 conversations over a month's time to establish a date for observation.

Over the Spring 2008 semester, five experts were recruited, three of which completed the study. For each subject, a similar process was followed: The intent of the study was described, the structure of the study (e.g., the purpose of the rig, the advise to take notes, etc.) was discussed, basic details of the cases were given, consent forms were provided, and then the study itself was performed. As mentioned, two subjects refused to sign the consent form and thus were excluded from the study; their reasons for refusal are discussed in Section 3. Each study was conducted at the officer's place of work, either a police station or a prison. Subjects were observed for up to two hours or until they felt that had made as much progress they could on the two cases. Officers were thanked for their time, and then the anonymized data was processed for initial analysis.

2.3 Study Post Mortem

Several observations from our subject data can be made. First, subjects did not make extensive use of the note taking capability of the forensics software or use the provided note paper regardless of expertise level. This frustrated our efforts to perform coding based upon these notes, leaving only the audio, video, and logging data. Secondly, the audio and video streams are quite noisy, possessing multiple interruptions of the task and non-task related questions by the experts. While useful information was mined from these streams, it was more labor intensive than initially planned. Data from logging provided more rich, and we have begun to code sequences of events using our analysis tool—i.e, identifying sequences of mouse clicks that correspond to search activity. Metrics such as click counts per subtask will be used to identify areas where the task performance could be improved and serve as candidates for visualization. Initial results indicate the searching for terms and their relationships cross documents are possible areas of improvement; formal and rigorous results from this analysis are beyond the scope of this paper.

3 Lessons Learned

In addition to the preliminary results from the study, we analyzed the process of performing the study. Given the length of time taken to elicit our three completed expert trials, we felt improvements could be made. Herein, we present the lessons we learned from this effort.

Keep the Goal in Mind The goal of our study is to observe how visualization may improve forensics. At several stages of our design, this eventual goal changed the nature of the study. As discussed, we initially considered a more thorough and intensive verbal analysis protocol. While this would be appropriate in the context of cognitive science, where the the low-level details of how a subject thinks is vital, for our goal more lightweight methods are sufficient. In addition, screen and mouse capture of the expert using the forensics tool was

added as some metrics for determining the complexity of a task cannot be measured without such logs. These metrics can then be compared to the same tasks performed using our eventual visualization solutions.

Working with Experts is Time Consuming While user studies in visualization are generally time consuming [9], expert populations require significant additional effort. Student populations for university-based studies are quite large, especially where Psychology programs provide subject pools as part of their curriculum. Experts, however, have to contend with their normal work assignments, which prolongs the process of recruitment and observation. Our forensic experts perform other tasks in addition to their forensic duties, complicating matters. Persistent effort (over 50 emails and phone calls were required for one subject, with 30 on average) is required. These factors confound the recruitment of experts, as demonstrated in the recruitment of software professionals for case studies [18]. We estimate that it took use two to three times longer to perform our initial study than it would have if we used only local, non-expert subjects.

Go to the Experts Though we had significant difficultly establishing contact with subjects, going to the experts provided valuable. First, it strengthens the relationship with the expert as it shows our willingness to work with them. Second, observations about the expert's work environment (such as the distractions during the study) are pertinent to understanding the user.

Clearly Communicate Expectations While this is good advice for any study, it is doubly important for experts. Our experts had no experience with human subject studies, so the goals and procedures were unclear to each. Part of the reason recruitment was protracted was due to anxiety over nature of the activity. One participant was concerned that the work would be used as part of their job performance evaluation (a false impression); another was unfamiliar with webmail cases, having dealt primarily with child pornography. With our later subjects, we were more clear with our expectations, and, as a result, the study went smoother.

Provide Consent Forms Early Though this ties in with the previous lesson, it deserves special mention. Though we communicated to all our potential subjects that they would be recorded and their interactions with the software logged, two of our recruited experts declined to participate when the consent forms for the study were presented. As the consent form is a binding agreement between the investigators and the experts, care must be taken in explaining the factors involved. In the case of the withdrawn experts, the consent form was rejected due to concerns about the study's data being subpoenaed at a later date as evidence of the expert's potential lack of proficiency; as required by Mississippi State Institutional Review Board policy (and stated on the consent form), data would have to turned over in such a circumstance. If we had provided the consent form during our initial contact with the expert, this issue would have been discovered sooner and other measures taken.

Be Prepared to Develop Your own Tools... During the design phase of our task study, we searched for software to assist in coding and analyzing the coded results. Our results we disappointing, and we found no off-the-shelf

software that would fit our needs. After queries to our empirical software engineering and cognitive science colleagues, we concluded that most such studies were conducted via spreadsheet software and labor intensive manual collating and coding. For our more lightweight approach (video, audio, and logging), we did not find any tools that made the analysis straightforward either. Thus, we ended up creating our own software for coordinating the video, audio, and logging events and for aggregating said events (Figure 1 right). In addition, we used several open source programs to assist in gather the data in the first place, though some commercial software exists for this purpose.

...But use the Tools the Experts Use An early decision of our group was to use the open source Autopsy software as our computer forensics platform; being open source, we could modify it to gather the logging data we required more readily than proprietary software. However, this proved to be a significant stumbling block with our experts, as they were in large familiar with the Forensic Toolkit (FTK). The lack of comfort with Autopsy uniformly caused extra training to be required before the experts could perform the study. For further iterations of our study, we plan to use FTK and have already instrumented it for this purpose.

4 Summary

Working with experts is required when an accurate understanding of their work practices is needed. We performed such a domain analysis to determine where visualization may benefit computer forensic practitioners. This study faces several unexpected hurdles which we have described as guidelines for visualization researchers interested in doing similar studies.

Though there were significant difficulties, working with experts was worth the effort. Our data has provided us with some initial avenues to pursue for visualization, and, more importantly, given us a better picture of how computer forensics is actually performed. We are currently redesigning our study to incorporate the atomic tasks we have identified such that novice users (i.e., university students) can perform them; this redesign is informed by our interaction with our experts. Finally, our pool of experts will be utilized to validate our visualization designs when they are complete. Such validation would prove more difficult without the groundwork of our initial study.

Acknowledgments

The work is funded by a National Science Foundation CyberTrust grant #CNS-0627407.

References

1. Householder, A., Houle, K., Dougherty, C.: Computer attack trends challenge internet security. IEEE Computer **35**(4) (2002) 5–7

2. Noblett, M., Pollit, M., Presley, L.: Recovering and examining computer forensic evidence. Forensic Science Communications **2**(4) (2000)
3. Wolfe, H.: Computer forensics. Computers and Security **22**(1) (2003) 26–28
4. Bequai, A.: Syndicated crime and international terrorism. Computers and Security **21**(4) (2002) 333–337
5. Kessler, G., Schirling, M.: Computer forensics: Cracking the books, cracking the case. Information Security (2002) 68–81
6. Thompson, R.: Chasing after 'petty' computer crime. IEEE Potentials **18**(1) (1999) 20–22
7. Teelink, S., Erbacher, R.F.: Improving the computer forensic analysis process through visualization. Communications of the ACM **49**(2) (2006) 71–75
8. Host, M., Runeson, P.: Checklists for sofware engineering case study research. In: International Symposium on Empirical Software Engineering and Measurement. (2007) 479–481
9. Kosara, R., Healey, C.G., Interrante, V., Laidlaw, D.H., Ware, C.: User studies: Why, how, and when? IEEE Computer Graphics and Applications **23**(4) (2003) 20–25
10. AccessData: (Forensic toolkit 2.0) http://www.accessdata.com/Products/ftk2test.aspx. Last checked May 2008.
11. Carrier, B.: (Autopsy forensic browser) http://www.sleuthkit.org/autopsy/. Last checked May 2008.
12. Carrier, B.: Ch. 11: Computer Forensics Basics. In: Know Your Enemy. 2nd edn. Addison Wesley (2004)
13. Singer, J., Lethbridge, T.: Methods for studying maintenance activities. In: Proceedings of the Workshop for Empirical Studies of Software Maintenance. (1996) 105–110
14. VanSomeren, M.W., Bernard, Y.F., Sandberg, J.A.C.: The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes. Academic Press (1994)
15. Hackos, J.T., Redish, J.C.: User and Task Analysis for Interface Design. John Wiley & Sons, Inc., New York (1998)
16. Hix, D., Hartson, H.R.: Developing User Interfaces: Ensuring Usability through Product & Process. John Wiley & Sons, Inc., New York (1993)
17. Mayhew, D.: The Usability Engineering Lifecycle: a Practitioner's Handbook for User Interface Design. Morgan Kaufmann Publishers, San Francisco (1999)
18. Sjoberg, D.I.K., Anda, B., Arishold, E., Dyba, T., Jorgensen, M., Karahasanovic, A., Koren, E.F., Vokac, M.: Conducting realistic experiments in software engineering. In: First International Symposium on Empirical Software Engineering. (2002) 17–26