Visualisation of Device Datasets to Assist Digital Forensic Investigation

Dr G Hales

Division of Computing and Maths,
School of Arts, Media and Computer Games
Abertay University
Dundee, UK
gavin.hales@abertay.ac.uk

Visualisation of Device Datasets to Assist Digital Forensic Investigation

Dr G Hales
Division of Computing and Maths,
School of Arts, Media and Computer Games
Abertay University
Dundee, UK
gavin.hales@abertay.ac.uk

***Abstract* - The increasing use of digital devices in our everyday lives, and their ever-increasing storage capacities places digital forensics investigatory resources under significant pressure. The workload for investigators is increasing, and the time required to analyse the datasets is not decreasing to compensate. This research looks at the potential for utilising information visualisation techniques to increase investigative efficiency with a view to decreasing the overall time taken to investigate a case, while still maintaining a high level of accuracy. It is envisaged that this may have the potential to lead to a reduced backlog of cases for law enforcement agencies, and expedited processing of criminal cases involving digital evidence.**

Keywords: digital forensics, information visualisation, computer security

I. INTRODUCTION

This paper aims to discuss the problems faced by digital forensics investigators, specifically those in law enforcement. There is potential for information visualisation techniques to be applied to datasets examined by digital forensics investigators with a view to increasing the efficiency of the investigator by aiding their comprehension of the data presented to them.

During the course of this research, a proof of concept tool, Insight, was developed to visualise results derived from the Autopsy 3, a mature open-source digital forensics application which provides substantial digital evidence analytic capabilities. The Insight software was used in an experiment to determine whether there were any benefits afforded by displaying information in a visual format, when compared to the textual format common in digital forensics tools such as Autopsy. The results of the experiment that was conducted to assess the potential gain in efficiency through the use of the Insight tool, and the use of exploratory information visualisation, will be explored in this paper and conclusions presented.

II. BACKGROUND

The area of digital forensics has long been an area in which investigators have been under constant pressure to keep pace with the workloads they are presented with. When computers started to become a common household item, law enforcement agencies found themselves having to deal with these devices as an important source of evidence. The initial response of government and law enforcement agencies to this new technology was to create organisations, such as the FBI's Computer Analysis and Response Team in 1984 [1], and methodologies to allow them to deal with this vast new source of evidence. Technology continued to advance at a rapid rate, with tool support in the area of digital forensics stagnating and leaving law enforcement agencies with sizable backlogs [2]. This is a problem which has persisted over the years as technology has continued to evolve. Often, a person will now own multiple devices which may include a laptop or PC, mobile phone and increasingly, a number of smart home devices such as the Amazon Alexa. All of these devices can be rich sources of evidence, and when coupled with the frequently large storage capacities of the devices, can generate a substantial workload for an investigator in a single case. Garfinkel (2010) argues that we are nearing the end of the 'Golden Age of Digital Forensics' due to increasing difficulties in processing data, both due to technical factors such as encryption, and also factors such as time-constraints. In a taxonomy of challenges faced in digital forensics, "vast volumes of data" is one of the significant challenges faced, along with "emerging technologies and devices" [3].

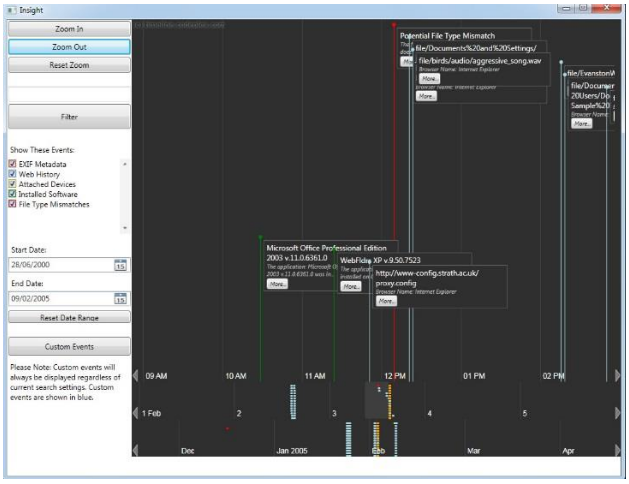
In most parts of a digital forensics investigation, the investigator is tasked with traversing large volumes of textual data which has been extracted from a device. This data can include the contents of documents, emails, file metadata, event logs etc. In looking through this data, the investigator largely has to make use of their intuition and experience with other cases to find relevant detail. As most of the information is presented to the investigator as text, it is very difficult for them to recognise patterns of behaviour or anomalies over a course of time. Generally, when a user is to be presented with large volumes of textual or numerical data, information visualisation techniques are utilised to aid interpretation of the data [4]. This spans as far back as 1786 when the Scottish engineer William Playfair invented visual methods of displaying economic data, such as bar graphs and line charts (Playfair, 1786). However, visual methods of displaying digital forensics data are largely unexplored, with many common digital forensics tools supporting either rudimentary information visualisation or no visual methods at all. This contrasts with the successful widespread use of visualisation in computer security software [5].

III. INSIGHT: VISUALISING DEVICE DATA

A. *Tool Development*

This research pursued the development of a tool which would visualise information from the Autopsy 3 forensics software. The Autopsy 3 tool was chosen because this is a popular open-source digital forensics tool providing substantial dataset analysis capabilities. However, the tool represents the majority of the acquired information in a textual format to the end user. As such, it was used as a pre-processor to pull information from a device image, and this information used as the foundation to build a visualisation tool on, called Insight.

As the Autopsy 3 software is a Windows-only solution, it was reasonable to develop the Insight software using the Microsoft .NET Framework in the C# language. A number of case studies were conducted to determine a suitable visualisation format. These studies examined different formats such as 2D and 3D, and different platforms such as web-based or desktop software. From these case studies, it was concluded that a tool which would provide the end user with a 2D timeline visualisation of the Autopsy dataset, and which would highlight various categories of events to the user would be the most suitable format to pursue (Figure 1).



Show These Events:
☑ EXIF Metadata
☑ Web History
☑ Attached Devices
☑ Installed Software
☑ File Type Mismatches

Start Date: 28/06/2000

End Date: 09/02/2005

Please Note: Custom events will always be displayed regardless of current search settings. Custom events are shown in blue.

Potential File Type Mismatch

file/Documents%20and%20Settings/

file/birds/audio/aggressive_song.wav
Browser Name: Internet Explorer
More..

Browser Name: Internet Explorer
More..

file/EvanstonW
file/Documer
20Users/Do
Sample%20
Browser Name:
More..

Microsoft Office Professional Edition 2003 v.11.0.6361.0
The application: Microsoft O
2003 v.11.0.6361.0 was in...
More..

WebFldrs XP v.9.50.7523
The applicat
installed on
More..

http://www-config.strath.ac.uk/proxy.config
Browser Name: Internet Explorer
More..

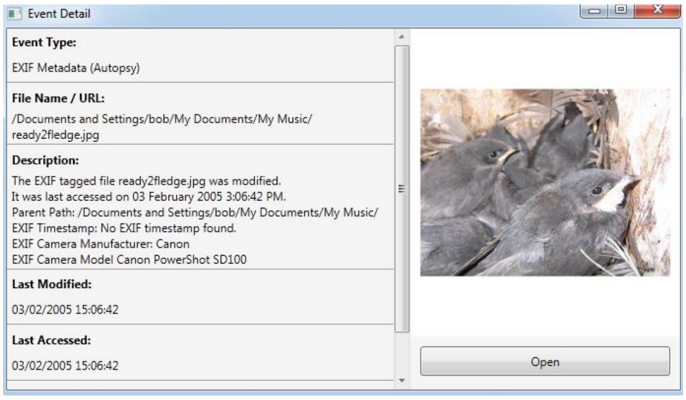
Figure 1 – Insight visualisation software

The developed software was designed to show the majority of the information available in the Autopsy software to the user in a more easily accessible format. In Autopsy, there are a number of different event categories the user can browse, such as EXIF metadata, web browsing history, software installations etc. As these categories are all displayed separately, it can be difficult for an investigator to derive narratives of user behaviour, as they have to keep navigating between categories. This leads to difficulty in correlating different types of data within the same time frame with each other.

The justification for utilising the timeline format of visualisation is that it is an exploratory visualisation format which presents the information from the case in a familiar chronological format. This may allow the investigator to view the information in a way which can assist them in recognising patterns and anomalies, and in recognising an overall narrative of user behaviour on the device. The software was also specifically designed to adhere to the Visual Information Seeking Mantra as defined by Shneiderman [6]. The mantra "Overview first, zoom and filter, details on demand" defines the ideal way in which visualisation software should operate to allow the user to navigate the information presented to them in an efficient and intuitive way. The Insight software does this by presenting all of the events from Autopsy in a colour coded format on a timeline. Overviews of the data are presented on two smaller timelines below the main timeline, both with increasingly larger time scales to the main timeline. This allows the investigator to navigate between points of time quickly and allows them to immediately see when there are bursts of user activity. The user can also zoom on the timelines to adjust the scale and give them a more detailed view of certain periods in time.

The software also provides the investigator with filtering facilities which allows them to select only certain types of events to display, and also to provide text to search for, thus allowing them to narrow their search. This can allow them to quickly filter out events which they know are not relevant to the investigation and give them a clearer view of the case.

Finally, the investigator has the option to click on individual events to open a detail window (Figure 2) which will give them more information. This window is also designed to allow the user to quickly view an image if that is what has been selected; this may be especially useful in a case where the suspect has been accused of possessing illegal images. This design satisfies the "details on demand" part of the Visual Information Seeking Mantra.



Event Type:
EXIF Metadata (Autopsy)

File Name / URL:
/Documents and Settings/bob/My Documents/My Music/ready2fledge.jpg

Description:
The EXIF tagged file ready2fledge.jpg was modified.
It was last accessed on 03 February 2005 3:06:42 PM.
Parent Path: /Documents and Settings/bob/My Documents/My Music/
EXIF Timestamp: No EXIF timestamp found.
EXIF Camera Manufacturer: Canon
EXIF Camera Model Canon PowerShot SD100

Last Modified:
03/02/2005 15:06:42

Last Accessed:
03/02/2005 15:06:42

Figure 2 – Event Detail Window

When looking at a case, often an investigator will have more than one device to examine; such is the case when the suspect owns a PC and a laptop, along with other digital devices. They are often required to examine these devices in isolation of each other, so it can be difficult to correlate information derived from one device with information on the other. For this reason, the functionality to add custom events to the timeline was included.

The reason for this is that it may be useful for the investigator to be able to create landmark points on the timeline to allow them to keep their bearings more easily, or to allow them to add known external data to the timeline. For example, the investigator may have evidence derived from another source which shows that the user was not in possession of the device between certain points in time. In this case, they can add custom points to the timeline which will make this clear to themselves and others involved in the case, allowing them to dismiss any events between these points as evidence.

B. Results

The software was tested with a group of trained Digital Forensics students (n=29) who were selected as they possessed the required baseline knowledge in order to complete a digital forensics investigation. The participants were given a set of 6 questions relating to a synthetic criminal case using a dataset created for teaching purposes. This dataset depicts the PC of a suspect, 'John Doe', who is accused of possessing contraband images. In this case, the participant is told that any image depicting a bird is to be considered contraband. The dataset is relatively small, around 5GB, and gives around 2 weeks of device usage on a Windows XP machine. Each of these questions given to participants asked them to find a different piece of evidence, or to draw a conclusion about the behaviour of the device owner. For example, one of the questions required that the user identify which brand and model of camera the suspect owned. This required them to use the tool to examine the EXIF metadata of the various images found on the device. Another question was more chronological in nature, and asked the investigator to identify where the suspect was on a specified day. In order to solve this, the investigator was required to explore all events on a specific day to find evidence that could link them to a location in the physical world. In this case, the suspect had visited the Wi-Fi login page of a university, so it can be assumed that they were on the university campus.

Of the 29 participants, 15 were given a copy of the Insight tool to investigate solutions to these tasks, and 14 participants were given a copy of the Autopsy 3 tool. They were instructed to complete the experiment with only the digital forensics tool allocated to them; and to time the experiment using a tool which had been developed to link the user's anonymous participant ID to the time they took to complete the experiment. The results were automatically uploaded to a remote server, thus removing potential inaccuracies of participants having to monitor their own time and report it back. The participants were asked not to discuss the case with other participants, and were required to complete the investigation in one sitting so as not to skew results.

When comparing the time taken to complete the experiment between the two participant groups, it was found there was no statistically significant difference between the groups ($\alpha = 0.05$). Although this does not indicate that there is an efficiency benefit to the investigator when using the visualisation software, it can also be viewed in a positive way; that is, when using the visualisation software there is no efficiency penalty, and the investigator can complete the investigation just as quickly. This means that any other benefits will not come at the cost of reduced overall efficiency.

Accuracy rates were also evaluated for participants, that is, whether the answers provided by the participant for each question was correct. The reason this was assessed was that if the Insight tool had been found to provide a significant improvement in investigation time, it was important to ensure that the participant was still providing answers that were correct [6]. In a criminal investigation, accuracy of investigators would be paramount so as not to provide incorrect conclusions, which could have severe consequences. It was found that in 1 of the 6 questions, there was a statistically significant improvement in accuracy rates for participants using the Insight software. This is promising as could indicate that with certain formats of question, Insight may provide benefits to the investigator. This is an area for future research.

Additionally, at the end of each question, the participant was asked to provide feedback on how easy they found each task to complete with the tool they had been given. This feedback included a Likert scale on which they were asked to indicate, on a scale of 1 – Very Difficult to 5 – Very Easy, how they found the task. Based on this feedback, it was found that in 2 of the 6 questions, participants responses were significantly ($\alpha = 0.05$) more positive for Insight than for Autopsy (Figure 3). This is interesting; as one of these questions (Q5) was the task discussed previously in which participant questions were significantly more accurate. By categorising the questions more clearly, it may be possible to establish whether certain types of question benefit from visualisation more than others.

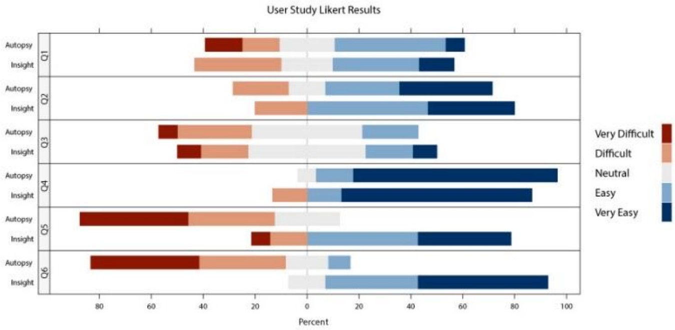


Figure 3 – Participant Task Feedback

IV. FUTURE RESEARCH DIRECTIONS

A number of benefits were identified in this research based on the empirical evidence gathered, in regards to investigative accuracy and ease of drawing a conclusion. However, these benefits were significant only in certain tasks the users were asked to perform. In future research, it would be of interest to classify the tasks based on what the user is being asked to do. This would allow for more detailed results to be gathered about whether visualisation aids the investigator in specific lines of investigation.

As part of this, it would also be beneficial to adjust the methodology so that participants are timed for each task instead of the entire investigation. This would allow results to be gathered which would reinforce whether visualisation provides a benefit for specific types of task. This would, however, require a significantly larger group of participants, or multiple different datasets, as it could be argued that the prior knowledge of the dataset from earlier tasks would artificially reduce the time taken to complete later tasks, thus introducing error.

V. CONCLUSION

This body of research has examined the various challenges currently faced by investigators in the field of digital forensics such as the ever-increasing difficulty in analysing large datasets across more devices. It has investigated the potential for information visualisation techniques to be applied in order to benefit the analytic capabilities of the investigator with a view to increasing the overall efficiency and accuracy of the digital forensic investigation process.

Empirical evidence gained from this research has shown that there is a potential for benefits to be realised through the use of exploratory timeline visualisations in terms of improved accuracy of conclusions derived from the dataset, and in some cases an improved user experience. Although, results from this research failed to show a statistically significant gain in investigative efficiency; that is, the time taken for participants to conduct an investigation was not significantly different when using the visualisation tool Insight to the time taken to complete an investigation using the frequently used Autopsy 3 software.

It is noted that the limitations of this research such as the synthetic nature of the dataset and of the tasks posed to the participants of the experiment may have influenced the time taken to complete the investigation. This is due to the fact that a real-life dataset could not be used in this experiment, as often the very nature of real datasets used in a digital forensics investigation is such that they are illegal to possess outside of a law enforcement context. A real dataset is also significantly larger than the dataset used in this research (around 5GB). This would have taken much longer to investigate, which was a constraint as the participants only had around 2 hours available to investigate the case provided to them.

Further research would be beneficial to reveal whether certain types of questions are common in a full-scale investigation. If certain types of questions are frequent enough and are similar in nature to those which showed a significant gain in accuracy or user experience in the experiment, there is a potential for a different result in terms of investigative efficiency as the investigator may be able to reach an accurate solution more rapidly.

REFERENCES

[1] M. Noblett, M. Pollitt and L. Presley, “Recovering and Examining Computer Forensic,” *Forensic Science Communications, FBI, 2(4),* 2000.

[2] E. Casey, M. Ferraro and L. Nguyen, “Investigation delayed is justice denied: Proposals for expediting forensic examinations of digital evidence,” *Journal of Forensic Sciences, 54(6),* p. 1353–1364, 2009.

[3] N. Karie and H. Venter, “Taxonomy of Challenges for Digital Forensics,” *Journal of Forensic Sciences,* pp. 885-893, 2015.

[4] J. Fekete, J. van Wijk, J. Stasko and C. North, “The Value of Information Visualization,” in *Information Visualisation, 4950,* Springer, 2008, pp. 1-18.

[5] G. Conti, Security data visualisation, San Francisco: No Starch Press, 2007.

[6] B. Shneiderman, “The eyes have it: a task by data type taxonomy for information,” *Proceedings 1996 IEEE Symposium on Visual Languages,* pp. 336-343, 1996.

[7] E. Casey, “A sea change in digital forensics and incident response,” *Digital Investigation,* vol. 17, pp. A1-A2, 2016.

[8] S. Garfinkel, “Digital forensics research: The next 10 years,” *Digital Investigation 7,* pp. S64-S73, 2010.

[9] W. Playfair, “The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century,” 1786.