

Assisting Digital Forensic Analysis via Exploratory Information Visualisation



A thesis submitted as partial fulfilment of the requirements of Abertay University for the
degree of Doctor of Philosophy

By

Mr Gavin Hales BSc (Hons)

School of Arts, Media, and Computer Games

February 2016

Declaration

Candidate's declarations:

I, Mr Gavin Hales, hereby certify that this thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy (PhD), Abertay University, is wholly my own work unless otherwise referenced or acknowledged. This work has not been submitted for any other qualification at any other academic institution.

Signed [candidates signature].....

Date.....

Supervisor's declaration:

I, Dr. R. I. Ferguson, hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy (PhD) in Abertay University and that the candidate is qualified to submit this thesis in application for that degree.

Signed [Principal Supervisors signature].....

Date.....

Certificate of Approval

I certify that this is a true and accurate version of the thesis approved by the examiners, and that all relevant ordinance regulations have been fulfilled.

Supervisor.....

Date.....

Acknowledgements

Many thanks to SICSA who provided the funding for this PhD. The resources provided by them have been of great use to this research.

I would like to thank my family for the unwavering support they have provided me over the years. From buying my very first programming book (Visual Basic 6), to providing advice and encouragement while I was frantically writing my thesis. My whole family have been so supportive, and for that I am eternally grateful. Without them, none of this would have been possible.

Thank you to Siobhan, my best friend of many years, who, even though we don't see each other as much as we may like, will always be my closest friend. Siobhan listened to many of my stressed out rants, and has always been a constant pillar of support to me. She managed to keep me sane through everything (well, relatively). I couldn't ask for a better friend.

Thanks to Ross, whose support was much appreciated in the later stages of my PhD. The panic of meeting my submission deadlines and sitting my viva was definitely eased by all of the food you made, and the gaming nights on the sofa! The ability to have a geeky rant with you will always be much appreciated too!

Thanks especially must go to Dr John Isaacs who has provided a wealth of opportunities and advice over the years, and who I will always consider to be a good friend.

Last but most certainly not least, many thanks to my PhD supervisors, Dr Ian Ferguson, and Dr Jackie Archibald. Your guidance, and support over the years has allowed me to reach this stage, and I am extremely grateful for this. Thanks also to all of the staff at Abertay who have provided PhD advice, and general support over the years.

Abstract

Background

Digital forensics is a rapidly expanding field, due to the continuing advances in computer technology and increases in data storage capabilities of devices. However, the tools supporting digital forensics investigations have not kept pace with this evolution, often leaving the investigator to analyse large volumes of textual data and rely heavily on their own intuition and experience.

Aim

This research proposes that given the ability of information visualisation to provide an end user with an intuitive way to rapidly analyse large volumes of complex data, such approaches could be applied to digital forensics datasets. Such methods will be investigated; supported by a review of literature regarding the use of such techniques in other fields. The hypothesis of this research body is that by utilising exploratory information visualisation techniques in the form of a tool to support digital forensic investigations, gains in investigative effectiveness can be realised.

Method

To test the hypothesis, this research examines three different case studies which look at different forms of information visualisation and their implementation with a digital forensic dataset. Two of these case studies take the form of prototype tools developed by the researcher, and one case study utilises a tool created by a third party research group. A pilot

study by the researcher is conducted on these cases, with the strengths and weaknesses of each being drawn into the next case study.

The culmination of these case studies is a prototype tool which was developed to resemble a timeline visualisation of the user behaviour on a device. This tool was subjected to an experiment involving a class of university digital forensics students who were given a number of questions about a synthetic digital forensic dataset. Approximately half were given the prototype tool, named Insight, to use, and the others given a common open-source tool. The assessed metrics included: how long the participants took to complete all tasks, how accurate their answers to the tasks were, and how easy the participants found the tasks to complete. They were also asked for their feedback at multiple points throughout the task.

Results

The results showed that there was a statistically significant increase in accuracy for one of the six tasks for the participants using the Insight prototype tool. Participants also found completing two of the six tasks significantly easier when using the prototype tool.

There were no statistically significant different difference between the completion times of both participant groups. There were no statistically significant differences in the accuracy of participant answers for five of the six tasks.

Conclusions

The results from this body of research show that there is evidence to suggest that there is the potential for gains in investigative effectiveness when information visualisation techniques are applied to a digital forensic dataset. Specifically, in some scenarios, the investigator can

draw conclusions which are more accurate than those drawn when using primarily textual tools. There is also evidence so suggest that the investigators found these conclusions to be reached significantly more easily when using a tool with a visual format. None of the scenarios led to the investigators being at a significant disadvantage in terms of accuracy or usability when using the prototype visual tool over the textual tool.

It is noted that this research did not show that the use of information visualisation techniques leads to any statistically significant difference in the time taken to complete a digital forensics investigation.

Table of Contents

Declaration	ii
Certificate of Approval	ii
Acknowledgements	iii
Abstract	iv
Background	iv
Aim	iv
Method	iv
Results	v
Conclusions	v
Table of Figures	xi
Index of Tables	xiii
Chapter 1 - Introduction	1
1.1 Background	1
1.2 Tool Support Gap	2
1.3 Proposed Solution	3
1.4 Research Question and Hypothesis	4
Chapter 2 - Literature Review	5
2.1 Introduction	5
2.2 History of Information Visualisation	6
2.3 History of Digital Forensics	18

2.4 Computer Security Visualisation Systems	32
2.5 Conclusion	41
Chapter 3 - Methodology	44
3.1 Introduction.....	44
3.2 Dataset Pre-Processing.....	47
3.3 Experimental Approach	51
3.4 Conclusion	56
Chapter 4 – Case Studies	57
4.1 Introduction.....	57
4.2 Case Study 1 - Timecubes - Pysight	57
4.2.1 Introduction.....	57
4.2.2 Development.....	57
4.2.3 The ‘Timecube’ model.....	59
4.2.4 Prototype Testing and Evaluation.....	61
4.2.5 Case Study Outcomes	64
4.3 Case Study 2 - Hyperbolic graphs - Walrus.....	66
4.3.1 Introduction.....	66
4.3.2 Prototype Testing and Evaluation.....	72
4.3.3 Case Study Outcomes	75
4.4 Case Study 3 - Timeline - Insight.js.....	77
4.4.1 Introduction.....	77

4.4.2 Development	78
4.4.2 Prototype Testing	81
4.4.3 Case Study Outcomes	83
4.5 Conclusion	86
Chapter 5 – Insight Prototype Development.....	88
5.1 Introduction.....	88
5.2 Insight	88
5.2.1 Prototype Development	88
5.2.2 Preliminary User Testing	98
5.2.3 Validatory Testing	100
5.3 Conclusion	105
Chapter 6 -Results.....	106
6.1 Introduction.....	106
6.2 Presentation of Results.....	107
6.2.1 Completion Time	107
6.2.2 Accuracy Rates	109
6.2.3 Likert Feedback Results.....	115
6.2.4 User Feedback.....	119
6.3 Conclusion	123
Chapter 7 - Discussion.....	125
7.1 Introduction.....	125

7.2 Discussion of Results	125
7.2.1 Completion Time	125
7.2.2 Accuracy Rate	128
7.2.3 Likert Scale Feedback	131
7.3 Research Outcomes	133
7.3.1 Validation of Research Hypothesis	133
7.3.2 Research Significance	135
7.3.3 Research Limitations	136
7.4 Further Work	138
7.5 Conclusion	142
Chapter 8 - Conclusion	143
References	146
Appendix A – Experiment Participant Sheets	152
Appendix B – Accuracy Results	159
Appendix C – Likert Feedback Results	161
Appendix D – Feedback Thematic Analysis	163
Appendix E – Participant Completion Times	166

Table of Figures

Figure 2.1- Causes of Mortality during the Crimean War	7
Figure 2.2 - World map showing countries by nominal GDP per capita in 2007.....	8
Figure 2.3 - Napoleon's Invasion of Russia	9
Figure 2.4 - Map of cholera infection in London (1854).....	10
Figure 2.5 - Exports of the UK (2012).....	11
Figure 2.6 - Hyperbolic models as shown by Munzner	12
Figure 2.7 - Urban Sustainability Visualisation.....	13
Figure 2.8 - “Church” vs “Beer” tweets across the US.....	15
Figure 2.9 - EMDialog.....	16
Figure 2.10 - History of Consumer Hard Drive Capacities - Wikimedia.org.....	23
Figure 2.11 - Autopsy 2 User Interface	29
Figure 2.12 - Autopsy 3 User Interface	29
Figure 2.13 – Webscavator	30
Figure 2.14 - Selection of views in Change-Link 2	31
Figure 2.15 - Spiceworks Dashboard.....	33
Figure 2.16 – ADO Log Visualisation.....	34
Figure 2.17 – RUMINT User Interface.....	35
Figure 2.18 - Snorby Dashboard.....	37
Figure 2.19 - Malware Binary Visualisation.....	38
Figure 2.20 - VERA visualisation of Netbull Virus protected with Mew Packer	39
Figure 3.1 - Autopsy SQLite database table structure	49
Figure 3.2 - Autopsy Ingest Modules	50
Figure 4.1 - Webscavator Timeline View (webscavator.org).....	60
Figure 4.2 - Timecube ‘layers’ representing days	61

Figure 4.3 – Side view of the time cube	63
Figure 4.4 - Snippet of LibSea graph file format	68
Figure 4.5 - Disk Image to Walrus Stages	69
Figure 4.6 - Windows 7 and XP clean installs in Walrus	70
Figure 4.7 - Walrus nodes coloured by file creation timestamp	72
Figure 4.8 – Insight.js Web Timeline	79
Figure 4.9 – Expanded event detail.....	79
Figure 4.10 - TimeGlider JSON Format	80
Figure 5.1 - .NET framework structure	90
Figure 5.2 – Insight showing 3 timelines with different time scales	94
Figure 5.3 – Insight event detail showing image file from source disk.....	95
Figure 5.4 – Insight filter panel.....	96
Figure 5.5 – Class Diagram	97
Figure 5.6 – Experiment Timer Application.....	103
Figure 6.1 – Comparison of accuracy rates between participants using Insight and Autopsy	110
Figure 6.2 – Likert-style feedback ratings for each question.....	117

Index of Tables

Table 6.1 – Completion Times.....	109
Table 6.2 - Accuracy of responses to each task in user study.....	110
Table 6.3 - Mann-Whitney test results.....	117
Table D.1 - Positive feedback themes for Insight.....	163
Table D.2 - Negative feedback themes for Insight	164
Table D.3 - Positive feedback themes for Autopsy 3	164
Table D.4 - Negative feedback themes for Autopsy 3	165

Chapter 1 - Introduction

1.1 Background

Digital devices are ubiquitous in our modern society, being almost entirely unavoidable, and increasingly being connected to the internet. People often have multiple electronic devices all of which are connected to the internet including desktop PCs, laptops, tablets, smartphones, games consoles etc; and increasingly, a number of other devices are being connected to the internet, which only a few years ago people would never have imagined would be connected to the internet. These devices include watches, televisions, smoke detectors, and even washing machines. As devices which all process human interaction, store data, and communicate with the outside world, these are often treasure troves of hidden information. Increasingly, one or more of these devices is involved when criminal activity is being conducted. This can include a drug dealer using his smartphone to interact with suppliers, a hacker breaking into a remote computer system using his laptop etc. As such, these devices are invaluable sources of information and evidence for law enforcement agencies.

The task of analysing these aforementioned devices is handled by specialists in the field of digital forensics. Similar to the way in which a traditional crime scene investigator would conduct a physical forensic search, these specialists analyse the data contained on the device in a forensically sound manner to find anything which may be of use to the case in question. This is a largely manual and often time consuming process. This is only made worse by the continuously rising storage capacities available in these devices, and the ever-decreasing costs. Digital forensics experts are therefore often found to be experiencing substantial backlogs, as they try to cope with the vast quantities of information they are required to analyse.

1.2 Tool Support Gap

A variety of different tools are available to support the digital forensics specialists at each stage of their investigation. However, even with this tool support, the investigators are still dealing with significant backlogs of data as they struggle to cope. The question must be asked as to why this is happening, and whether anything can be done to attempt to alleviate these problems. As will be discussed in this body of research, it is proposed that there exists a gap in this tool support. Many phases of the investigative process have good tool support, for example, the acquisition phase of the investigation is well supported. There exists a multitude of software applications and hardware devices which allow the investigator to create a forensically sound duplicate of the data contained on the device. However, when one examines the tools available to the investigator during the analysis stage of their investigation, in which they explore that dataset in an attempt to discover evidence and narratives of behaviour, it can be seen that these tools are not as well-developed. Many of the tools take the format of simplistic textual tools in which the user must manually search through all of the information. Although many of the tools provide features such as categorisation of information, and search features, the onus is still placed on the investigator to analyse the suspect dataset manually. The textual format of these tools means that they are not well suited to the rapid analysis of large datasets, as the investigator is required to read the displayed text line by line. It is proposed that in order to assist the investigator with their analysis of the data and to allow them to do this more efficiently and accurately, the format of these tools requires fundamental alteration.

1.3 Proposed Solution

This body of research proposes that in order to better assist the investigator with their analysis, the format of these tools should be shifted to one of a more visual nature. Information visualisation techniques have been shown in many cases to be an effective way of displaying large volumes of data to an end user, and to allow them to draw rapid conclusions from this dataset while removing the need to read and analyse every data point. The visual nature allows the end user to quickly recognise patterns and points of interest. It also affords them the ability to recognise ‘landmarks’ within the data, thus giving them an enhanced sense of context over the data.

This research aims to conduct a number of case studies which examine ways in which information visualisation techniques can be effectively applied to a digital forensics dataset. The lessons learnt from these case studies will be used to create a prototype information visualisation tool which displays information from the suspect device to the investigator in an easy to understand and intuitive way. The creation of this tool will allow experiments to be conducted which will determine whether the format of this tool provides a benefit to the end user, and if this could be a potential solution to the lack of modern and effective tool support for the analysis stages of a digital forensics investigation.

1.4 Research Question and Hypothesis

Having identified the gap in tool support for the analysis phase of digital forensics investigation, and the potential for information visualisation techniques to provide benefits for the investigator if applied correctly to a digital forensic dataset; the question posed by this is - “Does the implementation of information visualisation techniques, within digital forensics analysis phase support tools, provide the investigator with any benefits?”.

The research hypothesis can therefore be defined as: ‘the utilisation of exploratory information visualisation techniques in digital forensics investigations provides benefits in regards to effectiveness’. To define what is meant by effectiveness in this context; this can be broken down into two distinct metric for the purpose of this research. These metric include: the efficiency of the investigation, that is, how quickly the investigator can complete the investigation; and the accuracy of the investigation, that is, whether the conclusions drawn by the investigator are an accurate representation of the truth.

Chapter 2 - Literature Review

2.1 Introduction

This research will explore both the fields of Digital Forensics and Information Visualisation. This chapter will look at the history of these fields, their pioneers, and also relevant and significant research which has previously been conducted in this area.

Digital Forensics is a relatively young field of research, owing to it only coming into existence a number of years after the invention of the computer, and of the internet. The same, however, cannot be said of Information Visualisation. For many years, people have been displaying complex datasets as visual models. Computers have allowed these models to evolve into rich, interactive and complex entities; however, before this, visual models were created using pen and paper. Graphs and charts were drawn to convey complex information to stakeholders who were unlikely to understand raw data, and to provide people with information at a glance.

The evolution of these fields over the years, their relevance to this body of research, and any gaps which have been identified will be discussed in this section.

2.2 History of Information Visualisation

In the modern world, there are a multitude of complex and volatile systems, such as stock markets, computer networks and ecological monitoring systems which all generate vast volumes of data. This data comes in many formats; whether it be textual, numerical, images, videos etc. Often, the rate at which these complex systems generate data, or the format in which they generate it, means that it does not lend itself well to processing by humans. This data in its native format may be unwieldy in volume, or difficult to interpret, for example, a network firewall on the boundary of a large corporate file may generate hundreds or thousands of events per second, in a format which is highly technical. In an attempt to mitigate these difficulties, information visualisation systems have consistently been an important field of research (Schneiderman, 2008).

The aim of information visualisation is to take a dataset, and from this generate an accurate visual representation of the data. The purpose of this is to communicate the information in an easy to understand way, to allow the user a full overview of the data at one time, and to afford the viewer the opportunity to explore the data in a way that highlights patterns and relationships without obscuring the original underlying data itself.

Information visualisation techniques can be traced back to early pioneers such as the Scottish engineer William Playfair who invented visual methods of displaying economic data (Playfair, 1786). These included formats such as line graphs and bar charts in 1786, and the pie chart and circle graph in 1801. Florence Nightingale was another famous contributor to this area when she used a visualisation which is sometimes referred to as a Nightingale Rose to demonstrate the main causes of death during the Crimean War (Nightingale, 1858) (Figure

2.1). This is a famous example of when information visualisation has been used to convey information in an easy to understand format to important stakeholders. Nightingale used this to show that deaths caused by preventable illness far outnumbered the number of deaths caused by inflicted wounds. By using this simple to understand chart, Nightingale managed to successfully convince authorities in London, in the short time given to her, that improvements in hygiene standards in the hospitals would substantially reduce mortality rates.

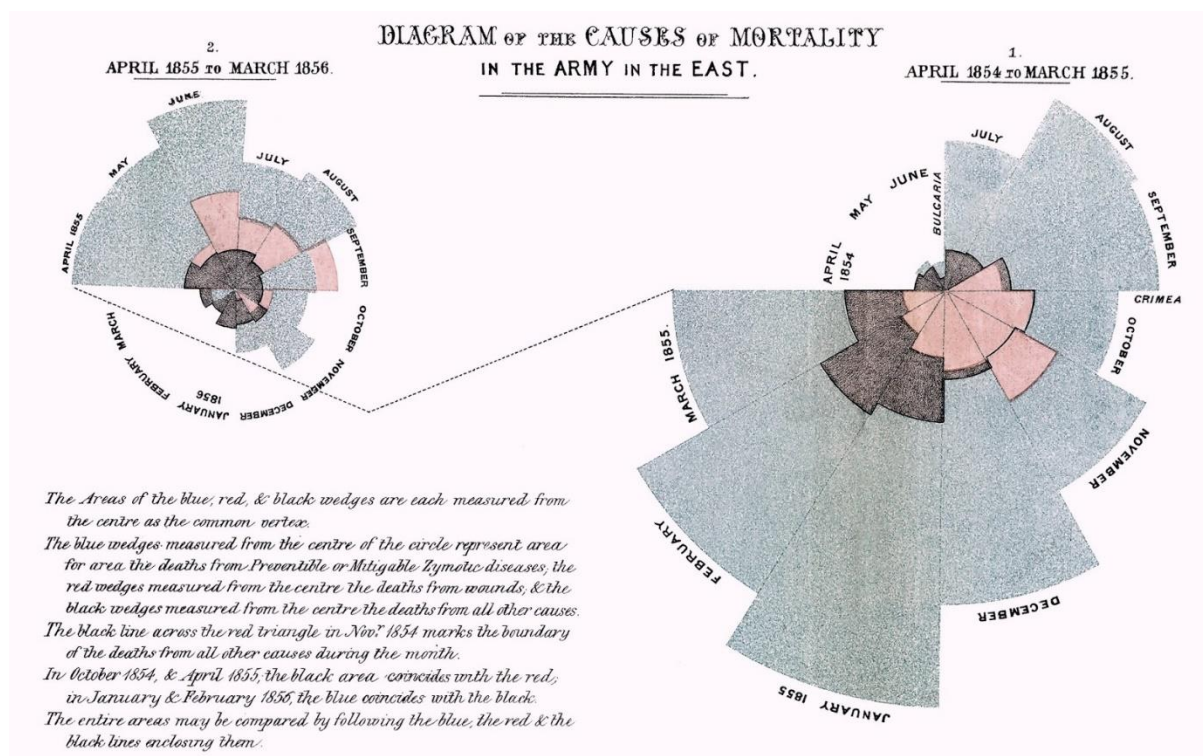


Figure 2.1- Causes of Mortality during the Crimean War

Geographical orientated information also lends itself well to visualisation methods. Often, numerical data may be shown by colouring areas of the world map to represent values such as GDP for each country (Figure 2.2 - World map showing countries by nominal GDP per capita in 2007 (Created By: Sbw01f, Wikimedia)).

Source: International Monetary Fund, as of April 2008

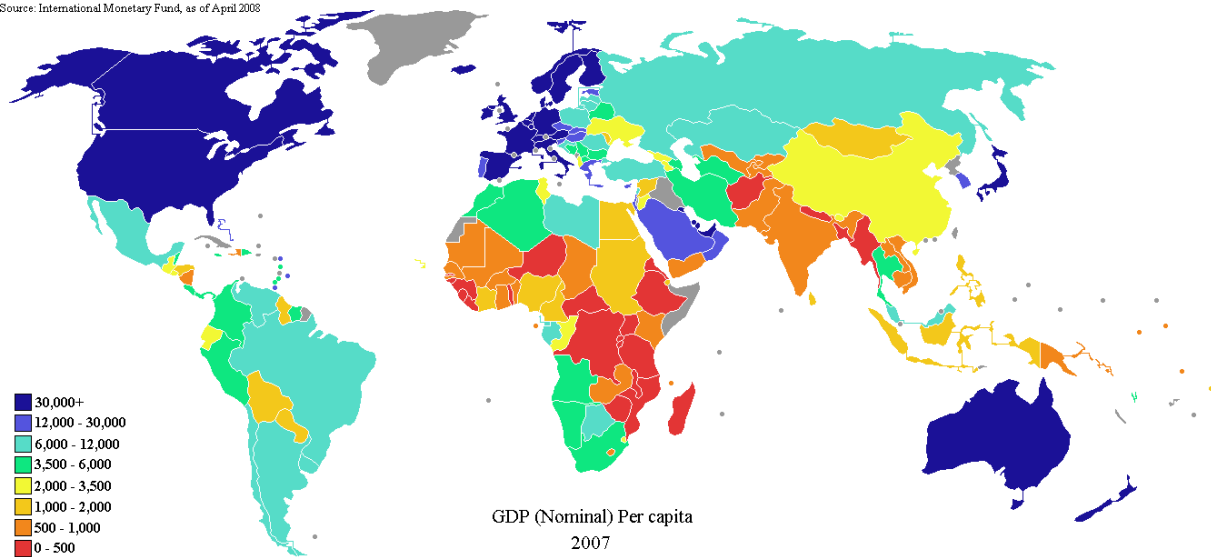


Figure 2.2 - World map showing countries by nominal GDP per capita in 2007 (Created By: Sbw01f, Wikimedia)

Arguably one of the most well-known early visualisations which maps a multifaceted dataset to a geographical representation is the map of Napoleon's invasion of Moscow in 1812 (Azzam, 2013) designed by Charles Joseph Minard (Figure 2.3). It can be thought of as one of the first Sankey (Sankey, 1898) diagrams. It not only shows the route Napoleon and his army took to Moscow, but it also visualises the size of the army at each point, the temperature, direction of travel etc. From this, it is possible to quickly see information such as how the army size decreased throughout the journey.

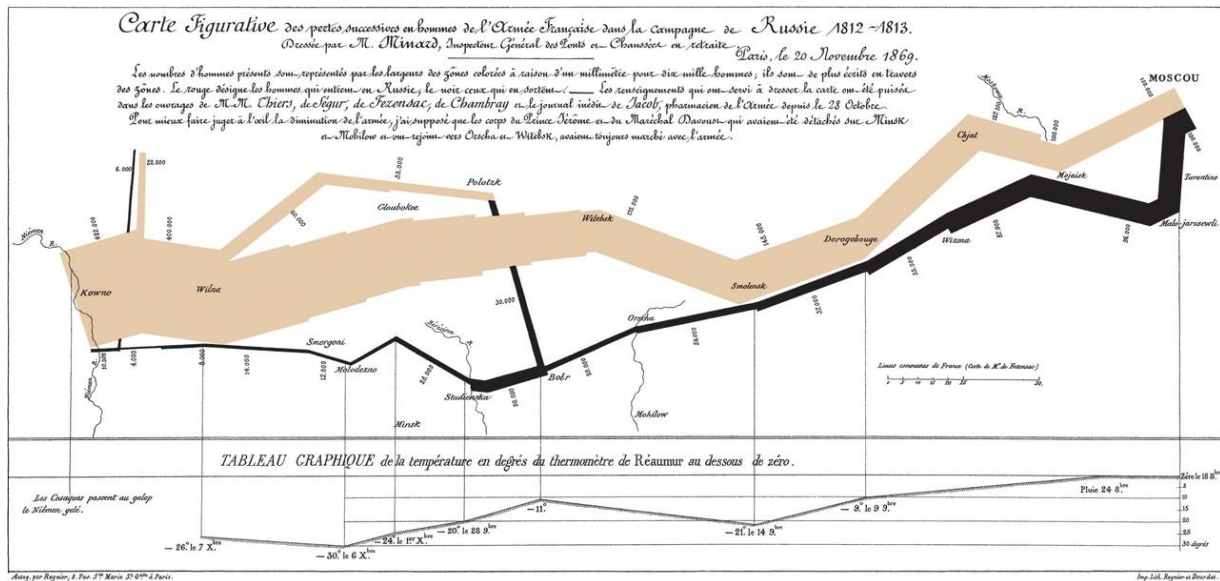


Figure 2.3 - Napoleon's Invasion of Russia

The value of information visualisation when attempting to find relationships in a dataset was demonstrated by Dr. John Snow (1855). Dr. Snow took the case of the cholera outbreak in Soho London in 1854, and used a dot map to mark the location each cluster of infection (Figure 2.4). As this was before the transmission method for cholera had been established, his map revealed a grouping of infection around the Broad Street water pump. From this observation, it was discovered that cholera is a waterborne disease and infected water had been fed into homes served by this pump, causing the outbreak. The visual model of infections assisted in this case by providing a visual link between the infection data, and the geographical locations of water pumps in the area. Had the infection data been analysed independently and without the aid of a visual model, it may have been very difficult to derive the relationship to the water pump.

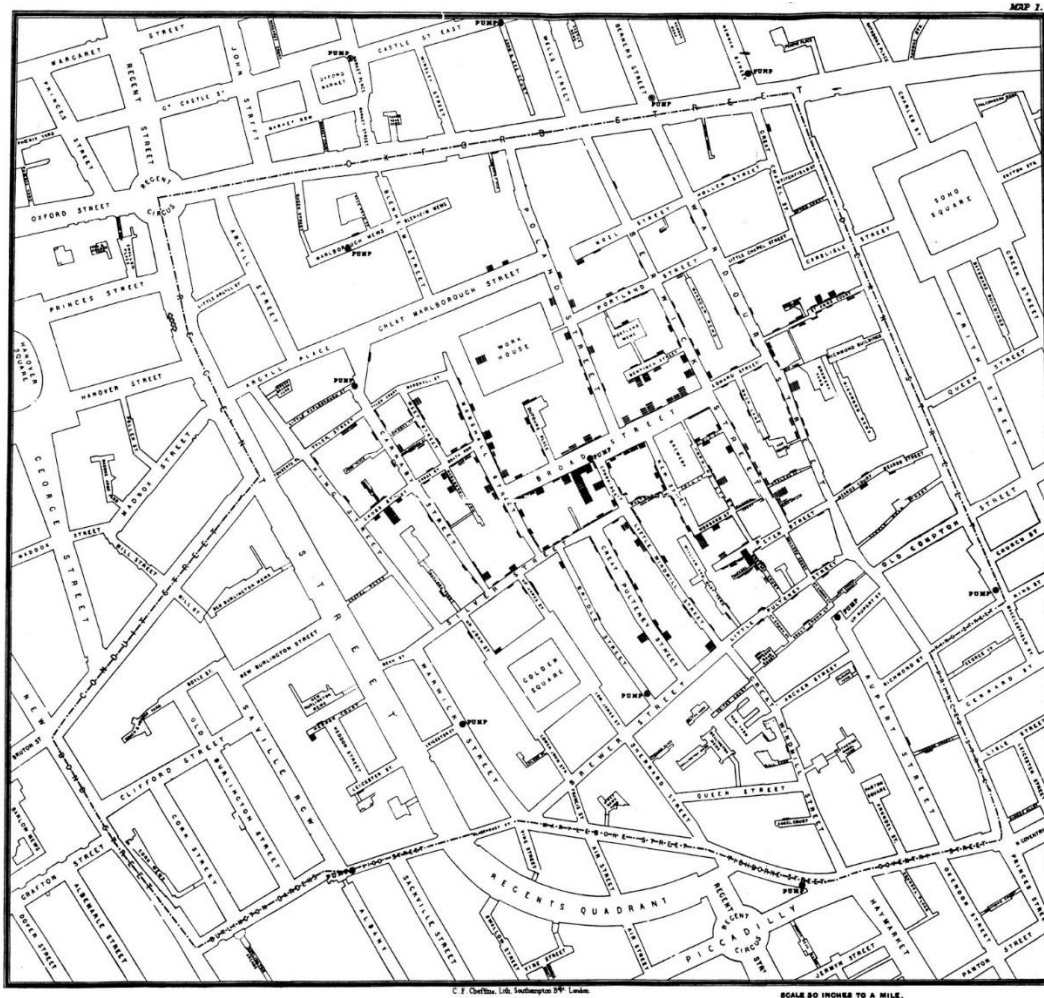


Figure 2.4 - Map of cholera infection in London (1854)

Approaches such as these were typical in early times in order to convey ideas and complex datasets to people who many not necessarily understand the raw data. The invention of computers and their ability to rapidly process data, and provide graphical feedback to the end user gave information visualisation experts a new medium with which to work.

One of the more modern pioneers of this research is Ben Shneiderman. His early work defined a new form of diagram to detail program flow in software applications. The Nassi-Shneiderman diagram (Nassi & Shneiderman, 1973) became a recognised alternative to the

standard flowchart. Later work by Schneiderman went on to describe the Treemap visualisation format (Shneiderman, 1992). This visualisation format was designed to show a hierarchy of data, which also included a quantitative element. As an example, this type of visualisation is often used to display data such as the exports of a country (Figure 2.5). By displaying the data in this format, the person viewing the diagram can quickly ascertain the primary exports of a country, and whether the exports of that country are quite diverse or generally involve a certain industry. If this data was displayed in a tabular format, it is difficult for the user to derive a sense of proportion between one export and another.

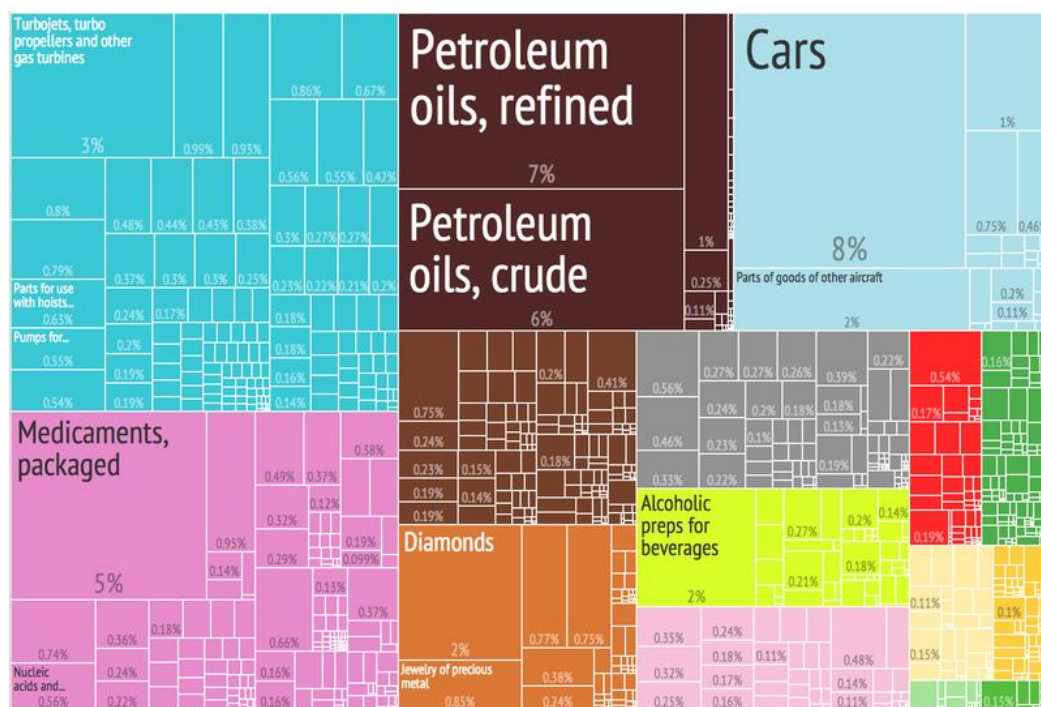


Figure 2.5 - Exports of the UK (2012)

Schneiderman also published a well-known paper, in which he defined the “Visual Information Seeking Mantra” (Shneiderman, 1996). In this paper, he defined the mantra “Overview first, zoom and filter, then details on-demand”. This mantra shows the steps which a user will generally take when using a visualisation application, and as such, should be kept

in mind when designing a visualisation application. First an overview of all of the data is shown to the user which will provide them with context, and may allow them to recognise obvious patterns in the entire dataset. They will then zoom to specific areas of the data, and use filtering tools to show only data that is of interest to them; allowing them to see how certain types of data may group together or otherwise be related. Finally the user can select individual data points, about which the application should provide in-depth details.

Around a similar time, Munzner (1995) conducted research into how hyperbolic models can be used to represent hierarchical data, such as the structure of the internet. Munzner's approach allows a large amount of hierarchical information to be displayed within a relatively small space (Figure 2.6).

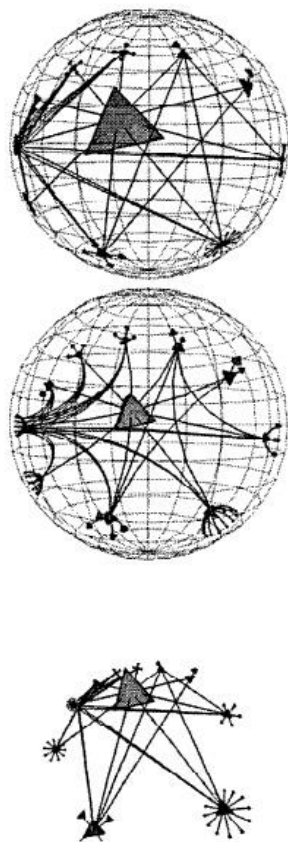


Figure 2.6 - Hyperbolic models as shown by Munzner

Visualisation techniques can also be used to map complex datasets to a visual model of a geographical area. Isaacs et al. (2011) demonstrated how a 3D model of a city can be visualised and coloured to show urban sustainability information. This information can include multiple attributes such as air emissions, economic output, energy efficiency etc. (Figure 2.7). By mapping multiple attributes to the model, the user can quickly assess the overall sustainability level for the building. As discussed by Isaacs, not only does this aid the understanding of the user; importantly, it also aids the understanding of any key stakeholders involved in the project. These stakeholders may not fully understand complex elements of the sustainability evaluation process. By visualising these elements, the stakeholders can interpret and understand important points quickly, and can participate more effectively in the project. Even in situations where the actual data is not overly complex, such as when planning the placement of a wind farm installation, visualisation tools can be used to assist the viewer in imagining what the effect of such an installation would be on the local landscape, and how it is predicted to change as a result (Bishop & Stock, 2010).

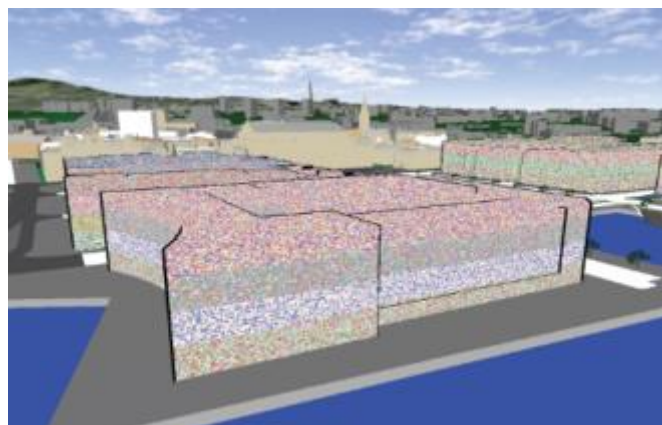


Figure 2.7 - Urban Sustainability Visualisation

The rising popularity of social networks such as Facebook and Twitter has led to the generation of vast amounts of data which can be analysed to provide insight into current affairs and trends. Often, these datasets can provide invaluable information, but the datasets are unwieldy to handle. Tools such as Apache Hadoop (White, 2012) for large batch datasets, and Apache Storm (Apache Storm, 2016) for real-time high-velocity datasets, can be utilised to rapidly process and generate information which is more manageable. This information can then be visualised to provide an effective overview of the data. For example, research has been conducted into sentiment analysis of tweets posted to Twitter, such as shown by Tan et al (2014) and Martínez-Cámara et al (2012). A simpler example of tweet analysis was created by a team of researchers at floatingsheep.org (Stephens, 2012) which analysed tweets across the US for the words “beer” or “church”. The outcome of this analysis was then visualised on a map of the US (Figure 2.8). From this it is possible to see large areas of the map such as the Southeast where religion is more prominent than others. The researchers made the interesting observation that a large number of the tweets in the areas with many more “church” tweets were generated by Foursquare check-ins at churches, in a way that they humorously described as “competitive churchgoing”. Other large datasets such as those taken from Facebook, or from research citation databases, can be analysed in a similar way and used to generate visual models of social networks (Henry, 2007). This can be of use to various organisations, for example, law enforcement agencies can use such information to build up knowledge of a criminal’s social network to hypothesise on relationships they may have with other criminals or groups of criminals. This may be of use in investigations where there are a number of suspects, whereby links may be formed between accomplices.

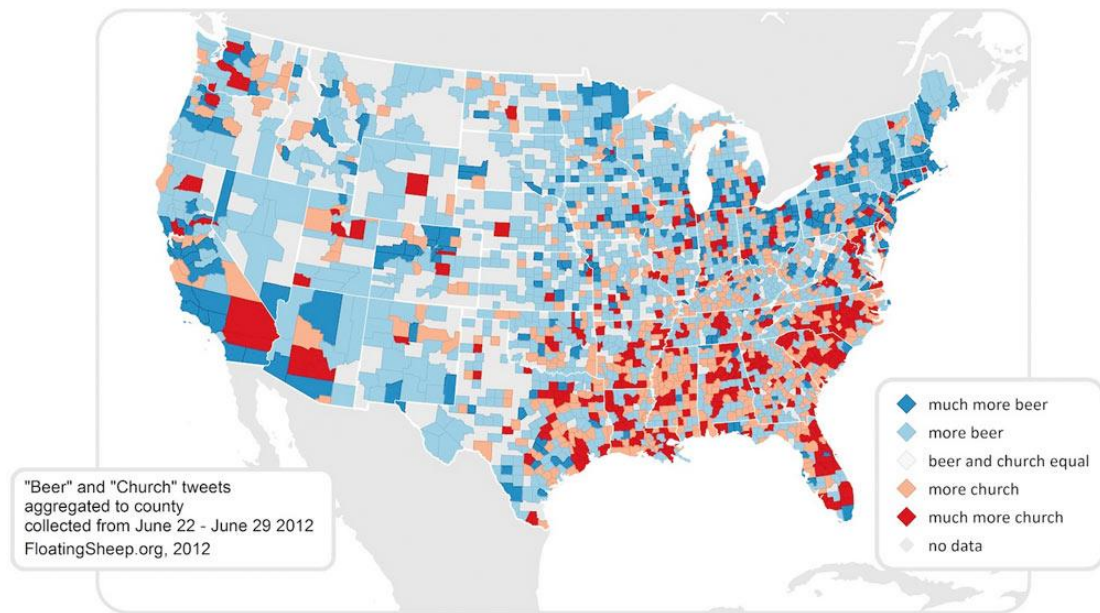


Figure 2.8 - “Church” vs “Beer” tweets across the US

It is not always the case that visualisation is used to make a complex dataset more accessible. Often, a relatively simple dataset can be visualised in a way that makes it more appealing, and simple to explore. As such, this can be a novel and effective way to introduce people to subject areas which are unfamiliar to them, allowing them to gain a basic understanding of the topic more quickly than if they were required to read a volume of text detailing the subject. Hinrichs (2008) presented such a system, EMDialog (Figure 2.9), which allowed visitors to the Glenbow Museum in Calgary, Canada, to explore information about the artist Emily Carr. The system gave the user two visualisation formats which they could use concurrently to explore the information; a time-based cut section diagram which is similar to the cross section of a tree, with each ring representing a decade of Emily’s life. This diagram is supplemented with a context based tree diagram, which provides links between snippets of information. This format allows the user to easily explore a vast amount of information on one screen, at their own pace; and as highlighted by Hinrichs, should reward the user with knowledge, regardless of whether they use the exhibit for a short or long period of time.



Figure 2.9 - EMDialog

It is the method of ‘exploratory’ information visualisation (Liu et al, 2013) that is of most interest to this research. Exploratory visualisation presents a large or complex dataset to the user in a visual manner and presents them with a meaningful and intuitive way to interact with the model. From this, the user is free to identify patterns in the dataset, and in general is free to draw their own conclusions.

As visualisation is becoming a more prominent research and development, numerous different toolkits and frameworks have been designed. Toolkits such as Prefuse (Heer et al, 2005), VTK, and D3.js all provide designers with APIs to quickly design visualisations. This allows for rapid prototyping and modification of visual models. One significant advantage that comes from this is that in a research context, for example, is that the researcher can

quickly create an early prototype visualisation tool, perform pilot studies and then quickly adapt their visualisation based on the outcomes of these studies. It also allows them to quickly design a number of different visualisation types, which can be tested against each other to find which is the most effective in that scenario. From a commercial point of view, this is advantageous as when a visualisation is being designed for a client, small changes which they may request will not necessitate significant changes to the software.

The rich history of information visualisation shows that throughout the years it has been successfully used to convey complex multi-faceted information in a way that is easy to understand, and to analyse. One key piece of work which shows the value of information visualisation when analysing information is that of Dr John Snow. By showing data points of cholera infections on a map, a pattern was shown which led to a medical breakthrough. It allowed him to see a pattern which may never have been visible through other means of analysis. It is this aspect of visualisation that drives its application in this research. By mapping digital forensics information in a novel way, it is envisioned that new ways to analyse this data may be realised.

2.3 History of Digital Forensics

Computing in its infancy was a convenience afforded to large corporations, academic institutions and governments. They would frequently own a single large mainframe which was operated by a number of people specifically trained to operate it. They were physically isolated from other appliances and had no real networking capabilities to speak of. It was Donn Parker, who in his book *Crime by Computer* (Parker, 1976) first noted the ability for evidence to be collected from these large appliances which had the potential to be used in court. Often, this was information which had been taken from audits which were conducted occasionally on the systems, and any discrepancies noted. In these instances, the number of suspects was very limited, as there was no real external access to the computer. There would also be a limited number of people in the organisation who had access to the system and who had appropriate training to be able to operate it. Large law enforcement agencies such as the FBI recognised this and trained some people in the basics of how to operate these complex machines who could assist other investigators when information was required from one of these mainframes. It was, however, at this stage not a priority in the slightest for these large organisations.

The field stayed relatively undeveloped until one of the early pioneers in the field of digital investigation noticed a minor 75 cent (USD) difference in two accounting applications that he was operating. Cliff Stoll, a systems administrator, documented in his book *The Cuckoo's Egg* (Stoll, 1989) his experience of using common tools available at the time to monitor the activities of hackers who were illegally accessing the system. By leaving a device connected to the line which the hacker had dialled in on, Stoll was able to watch the hacker's activity as they managed to break into the computer systems of a number of military facilities. After a while of tracking the hacker's activities and co-operating with various government

departments, the hacker was found to be one Markus Hess from Germany. Hess was found to have been selling information to the Soviet KGB.

By the early 1980s, PCs had been developed further and were becoming more readily accessible to the general consumer population, although were still at this time far from being what would today be considered as “user-friendly”. They were often purchased by hobbyists who would tinker with the internal workings of the computer and the code running on it. Pollitt (2010) notes that it was around this time that a few notable people from large US organisations, started to play a key role in the field of digital forensics. Pollitt recalls that these people included Mike Anderson, Danny Mares and Andy Fried from the Internal Revenue Service; Ron Peters and Jack Lewis from the Secret Service; Jim Christy and Karen Matthew from the Department of Defence; Tom Seipert, Roland Lascola and Sandy Mapstone from a few local U.S. law enforcement agencies. Two Canadians, Gord Hama and Steve Choy joined these people to form one of the earliest organisations dedicated solely to the area of digital forensics, the International Organisation on Computer Evidence (IOCE) (Whitcomb, 2002). In 1984, the FBI had established their in-house Computer Analysis and Response Team (CART) (Noblett et al., 2002), and other large US agencies developed their own groups, one of note being the U.S. Air Force Office of Special Investigation’s Computer Crime Investigator Program (CCI) which evolved into the well-known Defence Computer Forensic Laboratory (DCFL), the developers of the widely used free disk imaging software ‘dcfldd’. In 1993, the First International Conference on Computer Evidence was hosted by the FBI in Virginia with representatives from 26 countries (Pollitt, 2010). Pollitt also notes that the cases investigated at this stage mainly focused on data recovery due to the limited capacities of storage technology at the time. This meant that people were prone to deleting data frequently in an attempt to ensure they had free space available. The Internet was also

not in widespread use which meant that often there was little need to investigate anything more than one computer used by the suspect. Some criminals, however, chose to use dial-up internet access, and in some cases people found ways to hack the telephone systems to provide them with free calls etc. By doing this, they were able to hack remote targets for free, and could also manipulate the telephone system to disguise their identity. This practice was labelled as “phreaking” and was popular when telephone systems were primarily analogue systems (Collins, 1999). This has largely fallen out of use since many of the modern telecoms systems have been moved to digital methods of operation. However, “phreaking” techniques can still be used on older analogue systems which may still be found in developing countries.

Clearly this was not, however, solely an American problem. Many agencies around the world were starting to see problems which could arise from the growing criminal use of computers and started to also build their own groups dedicated to the field. In the UK, the Association of Chief Police Officers formed the Forensic Computing Group to try and deal with the rising threat of computer crime. The Computer Misuse Act 1990 was also brought into law in the UK after a well-publicised case in which hackers Steve Gold and Robert Schifreen managed to break into the personal BT Prestel mailbox of Prince Philip. The reason for the introduction of this law was that Gold and Schifreen were charged under the Forgery and Counterfeiting Act, and found guilty. However, the verdict was later overturned on appeal as the law was deemed to have been applied outside of its remit. The UK Government decided that a new law was required so that misuse of a computer was criminalized under an appropriate and relevant law.

Also during the early 90s, the tools used in digital forensics investigations started to take shape, evolving from the early command line tools being used for purposes for which they were never intended. A few key people in the field, many of whom were founding members of the IOCE, developed their own tools which were distributed and became widely used (Pollitt, 2010). These tools included examples such as IRS Utilities by Andy Fried, Maresware by Steve Ware. Although these were still command line tools, they were the first tools which were developed specifically for use in digital forensics investigations. The first commercial tool which targeted digital forensics investigations, SafeBack, was also released in 1991. It aimed to allow the investigator to easily create an exact duplicate of a drive.

In the period 1993 - 1995, the Internet became more frequently used in a domestic setting. This was due to the Mosaic web browser being launched and companies such as AOL beginning to sell dial-up access to the Internet. By 1995, Microsoft had launched Internet Explorer and there were around 19,000 websites available on the Internet which included Amazon and eBay (AVG, 2015). This rise in Internet activity, however, brought with it a stream of illegal use. A number of people had decided to use this new technology as a convenient way to share child pornography images. In an attempt to combat this, the FBI set up Operation Innocent Images in 1995 (FBI, 2006). Investigators would use chat rooms and bulletin boards frequently used by child sex offenders, and would pretend to be an underage child. When the predator arranged to meet the 'child', officers would be waiting for them at the location and they would be arrested. The FBI continued this program and as of 2006, the Innocent Images programme had led to 6,100 arrests being made.

It was around the early 2000s that Internet use started to rapidly accelerate. By this stage home computers were affordable and Internet packages sold by ISPs were readily available. In 1998, there were a total of around 1 million users of the Internet (AVG, 2015). In 2000 the first 512Kb consumer broadband service was launched in the UK, but it was relatively expensive in comparison to the widely used dial-up option. Over the next few years, the price of broadband connections dropped, while the speeds provided gradually rose, with the first 24Mb service being launched in 2005 (ThinkBroadband, 2014). In the same year, the number of global internet users hit the 1 billion mark. The capacities of hard drives were also increasing rapidly, which combined with the rising download speeds of internet connections meant that a user was able to consume and generate volumes of data never seen before (Figure 2.10). With drive capacities around 100GB in 2005, digital forensic investigators were facing an ever-increasing challenge. Recognising the need for tools to cope with the rapid change in technology, in 2002 a company called AccessData released their Encase Forensics tool (Dees, 2002). This tool was an 'all-in-one' solution which was designed to cover almost every stage of the investigative process, from imaging the suspect drive, to automatically providing analysis of the contents, to generating reports in a format which could be presented as evidence if required. This tool became the go-to piece of software for most digital forensic investigations, and a standard tool for law enforcement agencies.

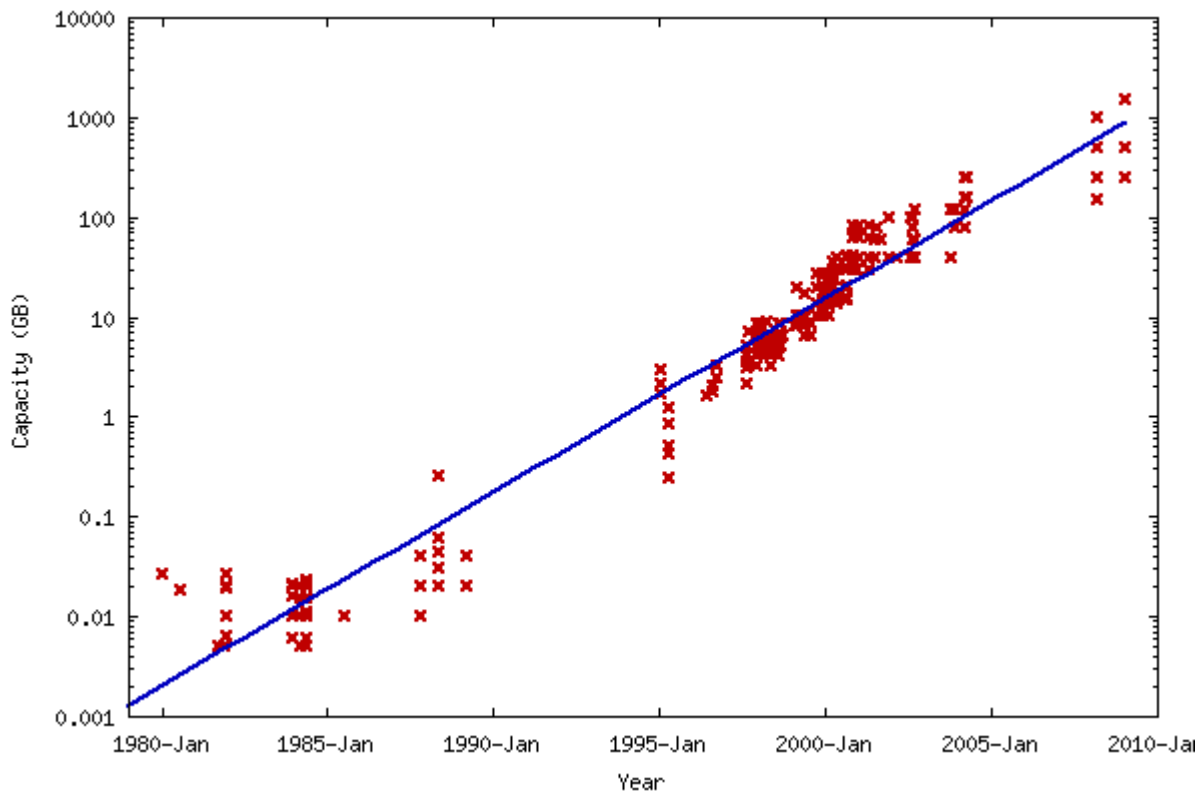


Figure 2.10- History of Consumer Hard Drive Capacities - Wikimedia.org

As shown in Figure 2.10, the volume of data held by consumer devices continues to rapidly increase, not only in respect to PCs, but also to the growing number of other complex digital devices which many people now own, such as smartphones. Even the smartwatch, a LG G Watch R, owned by this researcher has its own 4GB internal memory. Eoghan Casey, one of the leading voices in digital forensics research, has described the ubiquity of digital evidence on these devices as “digital dust”, leaving traces in generally unexpected places such as games consoles and sat-nav devices (Casey, 2010). As such, the volume of data which has to be processed in a digital forensic investigation has also increased, and now often comes from a wide variety of sources. Roussev et al (2013) note that in the period of 2003 - 2011, the volume of data processed per case on average increased by 6.65x, from 84 GB to 559GB. The authors focus on feasible methods to perform triage on the dataset, allowing the investigator to target areas of the data which may be valuable, instead of processing it in its

entirety. As Casey, Ferraro and Nguyen (2009) note, many Digital Forensic Laboratories have backlogs ranging between 6 to 12 months; utilising triage techniques could save valuable time by quickly identifying enough evidence to pursue a prosecution. Such methods may prevent a recurrence of cases such as U.S. vs. Brunette (U.S Government, 1999) in which the delay of the U.S. Government to examine one of the defendant's computers led to the evidence obtained from the device being suppressed in court. As the forensic analysis of the device began 2 days after the search warrant had expired, and no valid reason was given, only evidence taken from the first computer could be used. In this case the remaining evidence was admissible and the defendant was convicted. However, it highlights the problems that can be caused by delays in digital forensic investigations, and the possibility that due to these delays, a dangerous criminal could be allowed to walk free.

As the frequency of digital forensic investigations increases, a number of researchers have proposed methodologies in order to formalise the digital forensics investigation process. Such methodologies have been reviewed by Agarwal et al (2011) and Pollitt (2007), however, most proposed methodologies follow the basic forensic structure of Collection, Examination, Analysis and Reporting as defined by NIST (Kent et al, 2006). Proposed methodologies often expand this structure into stages of increasing granularity. Broadly speaking, the NIST defined stages, in a law-enforcement context, can be described as follows:

Collection: Devices which are relevant to the investigation are collected, recorded, and the data they contain is extracted and preserved in a forensically sound manner.

Examination: Digital forensics tools are used to extract relevant data from the devices, this can include file carving tools to recover deleted files from the media, and specialist applications and hardware to recover information from smartphones. Smartphones present a significant challenge at this stage, as often, the data is in a proprietary format; and in recent releases of major smartphone operating systems such as Android and iOS, encryption of the device is enabled by default. This makes the data almost impossible to access without the correct key from the device owner.

Analysis: At this stage, the data which is taken from the device is analysed by the investigator in an attempt to reconstruct narratives regarding the use of the device. The investigator will attempt to find evidence, whether inculpatory or exculpatory in nature, which is related to the accused crime. This analysis stage may also be directed to try and find pieces of evidence which may corroborate other non-digital evidence the investigator may already be in possession of.

Reporting: Information which has been discovered will be compiled into a report format which is suitable for presentation to non-experts, for example, in court. This report will either prove or disprove the alleged criminal act, and will detail the evidence which does so, in a way which hides much of the technical terminology unless it is absolutely necessary.

It is the Analysis stage that is of most concern to this research; as this stage is for the most part, largely reliant on the intuition and experience of the investigator. Teerlink and Erbacher (2006) acknowledged this problem of inadequate tool support in their research, and developed an early tool which visualised filesystem information as a treemap structure. Their

research showed that when compared to methods of analysing the filesystem using common Linux command line tools, when the visualisation software was used, the users were able to locate files on average 35% faster. This research shows that information visualisation has a high potential value if applied to this field.

Current tools, of which shall be discussed later in this section, typically present the investigator with a large volume of data, which they must manually navigate. Given the current storage capacities of devices, it is unreasonable to expect that an investigator should examine all of this data in depth. To do so would require a vast amount of time, which is not a luxury afforded to many investigators.

It should be noted that it is not only law-enforcement agencies who will conduct digital forensics investigations. It may be necessary for a company to make use of this type of an investigation for a variety of reasons. Should the company suffer from a hacking attack, a digital forensics investigation is often helpful to assess the extent of damage caused, if any (Garfinkel, 2013). Affected systems can be examined to discover traces of what has been done on the system such as whether any modifications were made which may present a continuing risk by creating security vulnerabilities. It is also important that the company can discover the nature of any information which may have been stolen from their systems. This is especially crucial for companies who are in possession of confidential customer information. The investigation should aim to rapidly determine the extent of any information theft; allowing the company to alert its customers in a timely manner.

There are now a variety of different tools (Pollitt, 2010) available to conduct digital forensic investigations. The tools which are used by large industries and law enforcement are typically expensive commercial software suites which are often described as “push-button forensics tools”. This means that once the investigator has secured a copy of the data held on a device, they can input this copy to the tool and it will automatically process the data, categorise it, and present it to the investigator, typically in a tabular format. Such tools do not remove the need for the investigator to thoroughly investigate the data, but simply process it into a format which is more accessible. Commercial tools include Encase by Guidance Software, mentioned previously, and FTK by AccessData. An open-source alternative to these tools, which is often used in a research context or by smaller companies, is Autopsy. This software provides functionality which is comparable to the commercial tools mentioned but is free-of-charge. As such, Autopsy (specifically Autopsy 3) will be used as a baseline for this type of tool throughout this research. As a result of the application also being open-source, it is also well-documented, which makes writing software which links into Autopsy much simpler.

These ‘push-button’ type tools often take much of the manual work out of a digital forensics investigation. For example, if separate tools were used, the investigator may have to collate log information from many different sources on the system, collect timestamp information from the file system, run file carving tools on the disk to attempt to recover any traces of files which had previously been stored on the system, but which the end user had subsequently deleted. This was very much the way in which early digital forensics investigations were conducted.

As discussed previously, digital forensics tools have not always been of the “push-button” type with powerful graphical user interfaces. Command-line tools have also existed which can perform much of the same Collection and Examination functions that modern GUI-based applications perform. They also exist for a wider range of operating systems, operating primarily on Linux but can also be used on Windows and Mac OS X. Examples of such open-source command line tools are The Sleuth Kit (Carrier, 2009), which allows for various kinds of disk analysis, and Foremost which attempts to recover deleted files from a disk by examining traces of data on the disk for common file type patterns. Due to the usability issues of these applications, a graphical front-end is often developed for users which will automatically interact with the command line tools. Autopsy 2 (Figure 2.11) is one such example of a graphical front-end. It is a locally-available web based system which uses The Sleuth Kit to retrieve information. The user is presented with a web page which allows them to browse the evidence in a tabular format. Autopsy 3 is an updated version of this software, which unlike Autopsy 2, was only designed to run on Windows PCs. It does not require separate command line tools, and is a fully integrated GUI application (Figure 2.12). Although aimed at consumers rather than digital forensics investigators, there are a number of tools which serve a similar purpose to Foremost; allowing consumers to retrieve deleted files from media. These pieces of software are often sold to serve one purpose such as to recover photo files from camera memory cards.

compromised, or that they were not the person to download these images. However, the web history may show that they have actively searched for this content, or visited sites known to distribute known contraband material. Other web history items such as sign-ins to Facebook may be able to also prove beyond reasonable doubt that it was the owner who was operating the device at this time. An example of software which was developed to provide such an insight into the web history of a PC is Webscavator (Lowman and Ferguson, 2011). Webscavator (Figure 2.13) analyses the browser history on a target device, and then displays the results in a number of different formats such as a timeline and word cloud. Such a tool is useful in allowing an investigator to discover the device owner's web browsing habits quickly, as the word cloud shows all of the information at a glance.

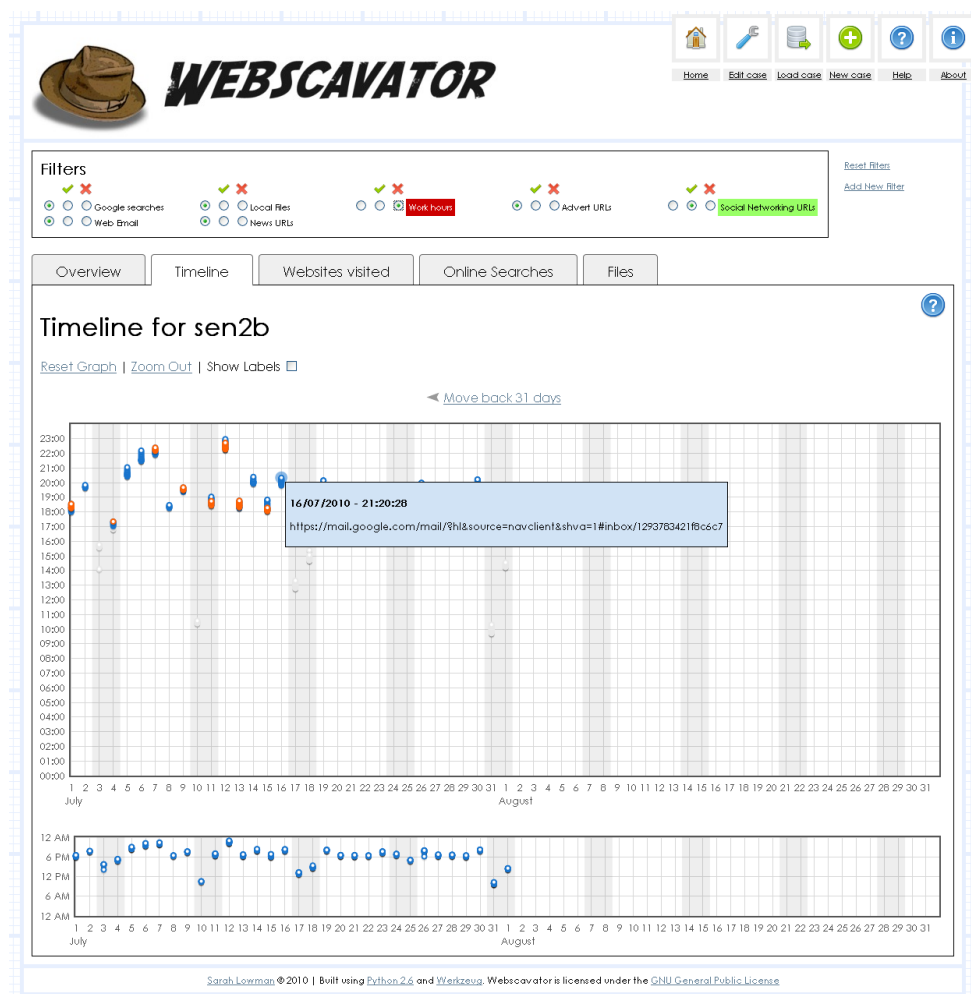


Figure 2.13 – Webscavator

Another piece of software which deals with a specific area of digital forensics is Change-Link 2 (Leschke and Nicholas, 2013). Many version of the Windows operating system have a feature called Shadow Copy which tracks changes to files, enabling features such as System Restore and Previous Versions which allow users to revert their system and/or files back to a previous state. Change-Link 2 (Figure 2.14) analyses these Shadow Copies, and provides the end user with a visual model of the changes. Such information is of great value to an investigator who may want to know what a user has done with their device since a certain point in time. It can be useful to determine when a user started to download large volumes of data to their device, or indeed, when large volumes of data may have been removed from the device. This can potentially be used to support or reject the alibi of a criminal suspect. For instance, if the user is charged with having downloaded contraband material, and they claim that they deleted the material as soon as they realised what it was, this can potentially disproved by examining Shadow Copies. If the file exists over a number of Shadow Copy snapshots, it can be said that the file in fact was not deleted quickly, but was present on the system for some length of time.

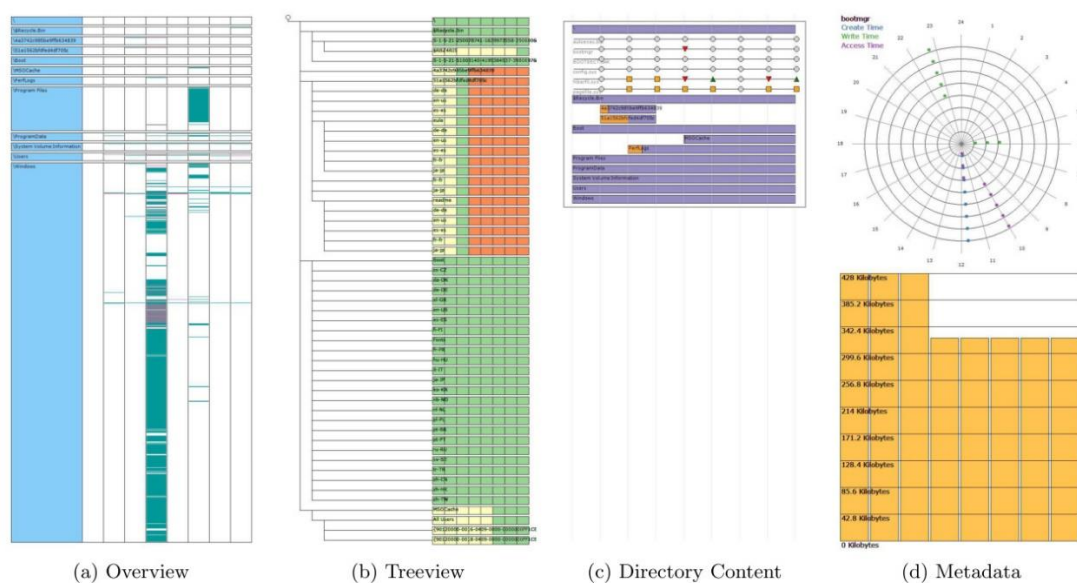


Figure 2.14 - Selection of views in Change-Link 2

The literature has shown that digital forensics is an area which is rapidly changing and evolving. Although tools are being produced which are attempting to keep pace with this evolution of the field, their improvements are in many cases limited to expedited acquisition of data from a source device. This research aims to examine ways in which the strengths of information visualisation can be applied to information gathered during a digital forensics investigation. Doing so has the potential to assist the end user in their analysis of the information, and improve the abilities of digital forensics tools to keep pace more effectively with the changes in the field. Early research, such as demonstrated by the Webscavator tool, demonstrate that such research is worthwhile pursuing. The timeline format of the tool is also of specific interest to this research, as it allows narratives of user behaviour to be drawn. This aids analysis of potentially criminal activity, as it can show whether the activity was persistent.

2.4 Computer Security Visualisation Systems

In the area of computer security, often datasets which are being dealt with are of a complex or high-velocity nature. This is especially true in the area of network security; a network administrator may be responsible for the security of a large corporate network, of which the manual analysis of all traffic flowing over the network would be an entirely unmanageable task. Even with security devices such as firewalls and intrusion detection systems, vast logs are often rapidly produced. As companies began to increasingly rely on their computer networks, and hacking became an increasingly publicised threat, the need for tools which would allow a network administrator to effectively deal with the volume of traffic was realised. Not only is the management of network traffic a problem, but the management of IT assets in a large enterprise is also a growing problem. An administrator will not only have to deal with the core infrastructure of the network but will also have to support the numerous

devices which may be assigned to a user. These can include devices such as desktop PCs, laptops, tablets and IP phones and even assets such as software licences. As such, there are now a number of commercial tools such as Spiceworks (2016), which allow the administrator to keep a track of the devices on their network, and quickly identify issues such as connectivity problems or out-of-date software (Figure 2.15). Simple visual methods are used in this type of software to quickly draw the administrator's attention to issues by highlighting them in red.

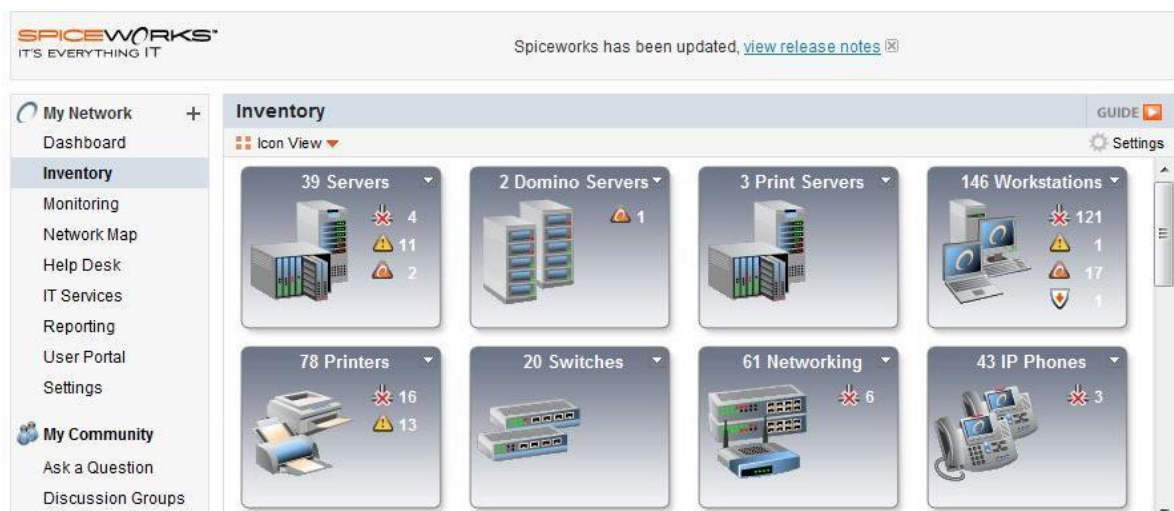


Figure 2.15 - Spiceworks Dashboard

A significant amount of research has been conducted into the area of network traffic visualisation. Schmerl et al (2010) demonstrate their ADO tool which shows how visualisation methods can be utilised to show relationships between different events generated by an Intrusion Detection System. Their research differs from many “dashboard” type tools, in the respect that it is an “explorative” type of visualisation. Many of the “dashboard” type of tools which show IDS data will give the user alerts about intrusions known to the system, or which are suspicious. ADO, however, allows the user to explore the

logs in a visual format and make their own decisions about what may be suspicious. The system will automatically cluster and link similar events, allowing the user to spot anomalous events quickly, and potentially link them to other events which may usually be missed (Figure 2.16). This is useful as the speed at which events occur on an IDS system preclude the ability for a user to manually analyse them, and even if they were to try, suspicious events would often be too spaced apart for the user to make any link. Schmerl et al make the example of an incorrect password being repeatedly attempted. The speed at which a person can type a password and make each attempt would give time for too many events to have occurred in the IDS logs for a user to even notice the break-in attempt happening. However, by clustering these attempts together on a visual model, the user can see them stand apart from other events, and can quickly investigate.

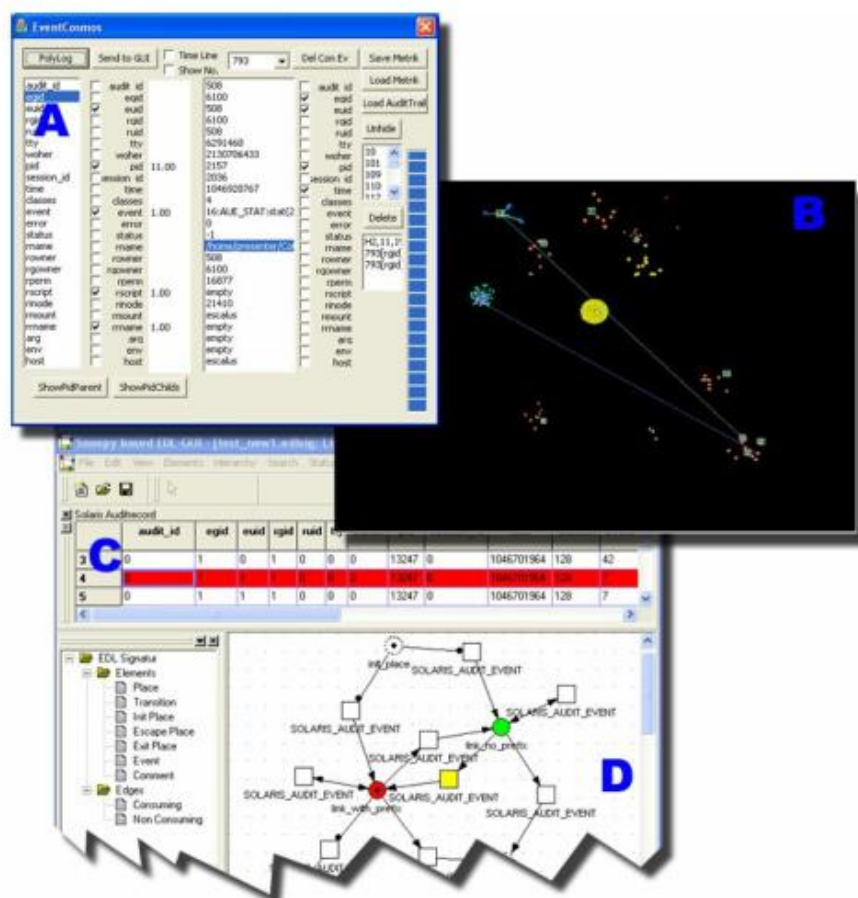


Figure 2.16 – ADO Log Visualisation

Another notable contributor to this area of research includes Conti, whose RUMINT tool allows the user to visualise Intrusion Detection System logs in a format which shows the information as a series of connecting lines, showing how traffic was flowing through the network (Conti, 2007). Source IP addresses are linked to source port which are then linked to destination ports and IP addresses. The user interface also provides a familiar VCR-like interface, allow the user to watch the traffic flow change (Figure 2.17). This functionality is useful as it makes anomalous traffic very apparent by allowing for familiarisation of the normal patterns the network traffic makes in the tool. It also gives the opportunity to analyse the state of the network in the moments leading up to any network attack. By doing this it may become apparent that there were in fact small indicators that a more complex attack on the network infrastructure was about to take place, for example, flows of traffic to uncommon port ranges.

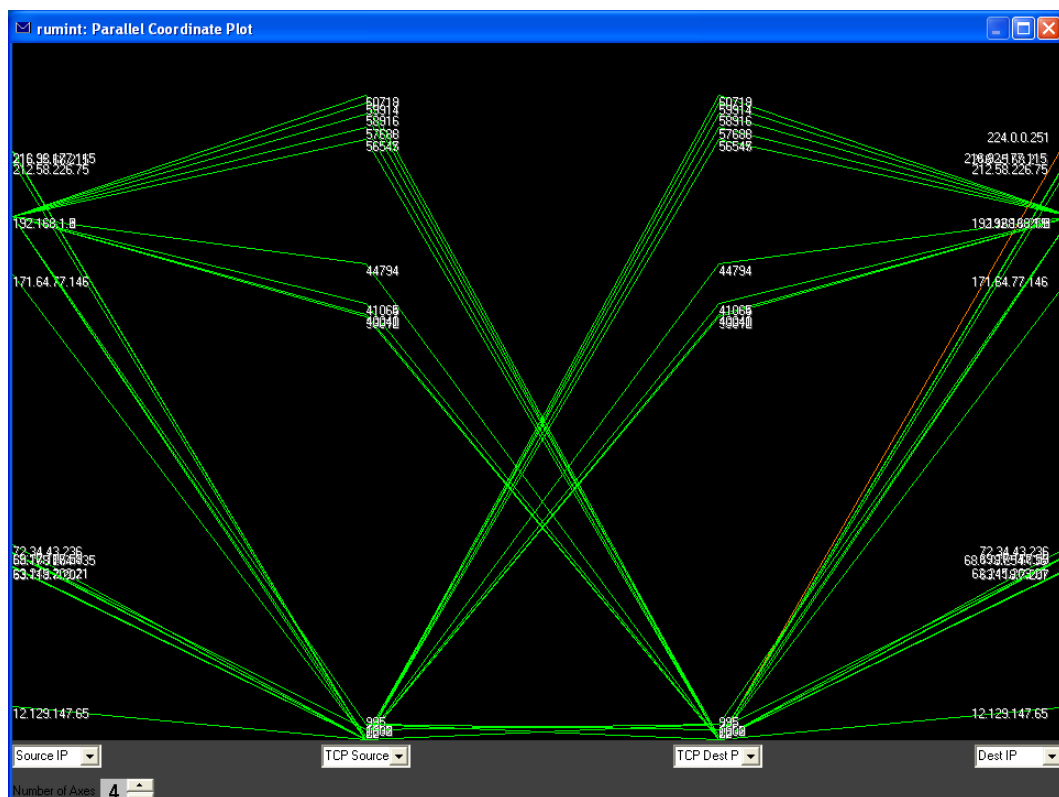


Figure 2.17 – RUMINT User Interface

Commercial and open-source software has also been developed to address this area. One such example is Snorby (Sanders & Smith, 2014). This is a web application which integrates with the Snort (Sanders & Smith, 2014) IDS software to provide the end user with traffic analysis capabilities. It utilises simple information visualisation techniques to show the user what is happening on their network; allowing them to make decisions to protect the network, and their data. This is especially important on a network where there may be devices attached which hold sensitive or business-critical data. The software automatically categorises the verbose log data generated by the Snort system and displays this on a graphical dashboard to the user. In doing this, the user can gain an understanding of what the typical use of their network looks like, and allows them to rapidly spot anomalies which may manifest as abnormal protocol types, or may even be simply flagged by Snorby as intrusion traffic. This type of traffic is highly visible in Snorby, as red “High Severity” events shown on the dashboard (Figure 2.18). This allows them to take action to protect their data, or to ensure the continuity of their business operations. Without using visualisation techniques, the manual analysis of the IDS logs would be a laborious process, and in many cases, this would mean that any intrusions would not be noticed until long after they had happened. This would be far too late for the administrator to take preventative action, meaning data loss may have already occurred. This is very important, as theft of digital data from large companies is becoming more widespread.



Figure 2.18 - Snorby Dashboard

Not only is hacking becoming a more widespread and publicised activity, but with the vast number of devices that are online, the spread of malware is also becoming a serious problem. Recent years have seen malware such as the now infamous Cryptolocker (McDermott, 2015). This one piece of malware wrought havoc on computer systems of personal and commercial users alike. Once on the system, it proceeded to employ public key encryption methods to effectively lock the user out of their own files and hold them to ransom; demanding payment in order to release the encryption key which would reverse the process and allow the user access to their files. The effect of this was not only the loss of sentimental and important files for personal users; but many businesses, especially small businesses, were locked out of important files crucial to their business. This led to lost income as time had to be taken to restore files from backups, and in some cases the businesses had no backups and lost access

entirely. Some users resorted to paying the fee, thus funding criminal activities and even then were not guaranteed access to their files.

This spread of malware and the associated losses caused by it have drawn significant research and commercial interest. This has led to a number of interesting attempts to implement visualisation techniques to assist in the resistance against the spread of malware. Such research was conducted by Nataraj et al. (2011) who suggested the novel technique of visualising the binaries of malware executables. Their approach to this was to take the binary data of the file, convert this to a vector of 8 bit unsigned integers and then to finally convert this into a grayscale image (Figure 2.19). This research is an early approach to visualising malware, but still it shows promising results. Often, the authors were able to demonstrate that variants of malware from the same family could be recognised from the visual representations. However, they note that there are a number of caveats to this method, such as the packing tools which are used to compress the binary and malware which utilises a polymorphic engine. Such instances will generate a drastically different visual representation.

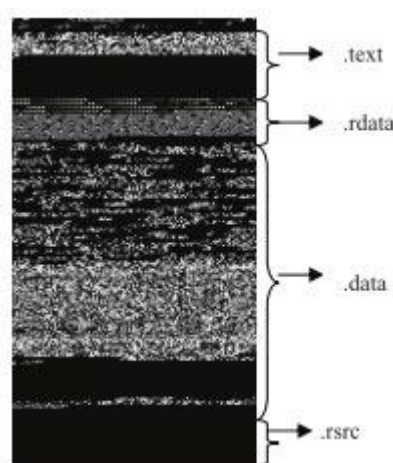


Figure 2.19 - Malware Binary Visualisation

Quist and Liebrock (2011) decided to take a slightly different approach to visualising malware. Their research involved the development of a tool, VERA, which provide a visual representation of the execution path of a malicious executable. This was done by first disassembling the malware binary using common monitoring tools and debugging software. The results from this were then used in the VERA (Figure 2.20) software to provide the end user with a visual model which represents the execution path of the malware. This leads to certain code paths creating recognisable patterns, such as loops and branches. From this it can be used to recognise specific strains of malware, and even to identify features common to different pieces of malware. It can be conjectured that by using this approach, it may even be possible to identify different pieces of malware written by the same person. This may be possible, for example, if the malware author has written a code library which they then use in the different pieces of malware they write. This section of code would presumably create a recognisable visual pattern, which would be identifiable in the associated binaries. This approach attempts to combat the problem of executable packing by deconstructing it and visualising it as part of the execution process.

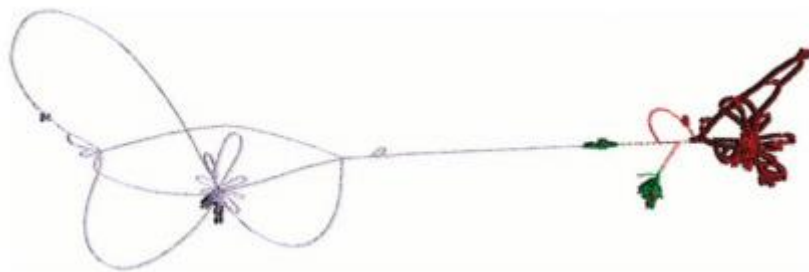


Figure 2.20 - VERA visualisation of Netbull Virus protected with Mew Packer

When reviewing the literature regarding the application of information visualisation to the area of computer security, it can be seen that this has been highly successful, as demonstrated by its use in many well-known commercial tools. From a research perspective, the use of visualisation to map the behaviour and structure of various forms of malware is of specific interest. As the aim of a digital forensics investigation is often to construct a narrative of user behaviour through the actions they have taken on a device, it can be said this is not entirely dissimilar to the approach taken to visualising malware behaviour. However, rather than mapping the processes of malicious software, this research aims to map the complex behaviours of the device user. As digital forensics is very closely related to computer security by the nature of the data involved, it is hypothesised that these visualisation techniques applied to computer security could be adapted to suit a digital forensics investigation, and could be equally as effective.

2.5 Conclusion

The review of available literature has shown that information visualisation is by no means a modern concept. In fact, it has been applied in various forms over the years, and the growing popularity and abilities of computer systems has provided a new way to process and present these visual models. Its successful application in many different fields is the driving force behind this research. Wherever there is a complex dataset which must be manually processed by a human, there is always an opportunity to ask the question: “How can we present this data in a different way, which can aid our understanding?”

Humans are naturally able to process visual information much more rapidly (Fekete et al., 2008) than we can read and process textual information. We are able to draw conclusions and relationships with just a small amount of time viewing visual data, and we can often memorise this visual information well, allowing us to recognise patterns which we may have encountered before, or seen elsewhere in the dataset. It is these traits that are exploited by the wide variety of information visualisation applications which were discussed in the previous section. Information visualisation also provides a way in which a very large dataset can be viewed in a limited amount of physical space.

Digital forensics investigations look to analyse patterns in information presented by the device undergoing analysis. The literature shows that one of the early examples of an application of information techniques in assisting this ability to recognise patterns is the work of Dr John Snow. As a researcher presented with a dataset of instances of cholera infection in London. When Dr Snow took this information and visually marked each instance on a map, he was able to spot a pattern of cases clustering around areas where water pumps were

located. The discovery that cholera is a water-borne illness could have been significantly delayed had this approach of visualising the information not been taken. The clustering of the data points led Dr Snow to examine what was located in the close vicinity of these areas, and to look further than what the raw data could tell him. It is this approach that this body of research aims to examine. Can the data from a digital forensics investigation be visualised in such a way that it can assist the examiner in looking beyond what the raw data shows; to allow them to synthesise higher level profiles of user behaviour, and perhaps even provide insight into their motives at certain points in time?

The literature regarding computer security is also supportive of the use of information visualisation in this field. Schmerl et al (2010) discuss the high velocity of data sourced from a network intrusion detection device. In a large network, the huge amount of data generated by these devices is so large, that if analysed manually, could mask real malicious events. They use the example of a person trying to manually guess the password of an account. The number of events between each attempt may be so great that it is impossible to spot these reoccurring break-in attempts. Their ADO tool is designed to allow the end user to visualise this information and to spot events which are outliers from what is normal. Quist and Liebrock (2011) also demonstrate an approach to visualising the execution path of malware samples. This is especially interesting as it provides a method for the “behaviour” of this malware to be visually represented. This idea of allowing recognition of patterns of behaviour is what this research aims to achieve when applying similar approaches to digital forensics investigations. Instead of looking at data generated by the behaviours of automated processes, the data generated by human factors will be examined. Given the paucity in the ability of digital forensics tools to assist in the analysis phase of digital forensic investigation, and the ever-increasing workloads placed on organisations such as law enforcement (Casey, Ferraro

and Nguyen, 2009), it is of utmost importance that steps be taken to improve the tools available to such organisations.

In summary, the literature has shown that information has a rich history which has demonstrated the ability to improve human analysis in a variety of different subject areas. The literature has also identified that the current tools utilised in digital forensics investigations are lacking in their ability to allow digital forensic examiners to keep pace with their continually increasing workloads. However, it is also noted that information visualisation techniques have rarely been applied to this field. In cases where such techniques have been applied to a sub-set of the information, such as web browsing history (Lowman and Ferguson, 2011), the results have been promising. As such, it is the objective of this research to examine the various information visualisation techniques which could be applied to digital forensics datasets, and to test the hypothesis that the application of an appropriate visualisation technique can realise gains in effectiveness for the investigator.

Chapter 3 - Methodology

3.1 Introduction

In the previous chapter, existing research and solutions in the areas of digital forensics were discussed. It can be seen from this examination that there is a gap in the development of digital forensics tools, specifically those that support the analysis phase of the investigative process. The benefits of information visualisation were also discussed, along with the variety of modern frameworks available to support this area. From this, it can be seen that given the primarily textual nature of digital forensics data, and its sheer volume, that a synergy may exist between these two areas, which may be possible to exploit in order to assist investigations and improve their overall effectiveness.

In order to examine the potential of the application of information visualisation techniques to this area, it was proposed that a methodology be followed in which multiple case studies were developed and examined, the benefits and pitfalls found in each, and the lessons learned, fed into the next case study. The desired result from this was to develop a framework and prototype tool which would allow experiments to be conducted with the aim of assessing whether there are any benefits afforded by utilising information visualisation techniques in digital forensics investigations. These case studies and final prototype tool would give an indication as to whether information visualisation is a viable option for bridging the tool gap which was identified in the Literature Review chapter. They would also identify any areas for potential further research.

There is, however, a distinct problem in the area of digital forensics regarding the availability of datasets for use in testing. It is not possible to gain access to realistic datasets for a number

of reasons. As the test datasets which are used are full copies of computer hard drives, these will almost always include personal or highly sensitive information. This can be partially mitigated by using disk images from people who have provided consent, with the understanding that their data will not be published in its raw format, and that only this researcher will have access to the data. Any information which was to be published would be heavily redacted, however, this also poses problems in that it may not be possible to redact the information in a way that completely protects the identity of the donor, without significantly compromising the usefulness of the information. Also, given the volume of information held on a disk image, it could be very easy to accidentally miss something. Any disk images which were donated were stored on a fully encrypted drive to protect them from theft or loss.

Digital forensics investigations are also usually conducted to gather evidence about a criminal act. The nature of this information, more often than not, also means that its possession alone is a criminal act. One of the prime examples of this is child pornography, one of the types of crime most often dealt with by law enforcement digital forensics units. As such, it is impossible to test this research using a dataset which reflects a real world use case.

In order to test each case study, a synthetic dataset was used. In this dataset, it assumed that any information of an ornithological nature is classed as contraband material. This includes images of birds, and email exchanges regarding the subject etc. This dataset has been developed by Dr Ferguson over ten years teaching of the digital forensic subject area, and provides an example scenario which contains data spanning a few weeks. This dataset will be referred to throughout this research as “The John Doe Test Case”. Although this dataset is quite small, it allows us to quickly test various tools with a dataset which provides a known

narrative of activity. The disadvantage to this dataset is the volume of data it contains (around 5 GB) Although when this dataset was created, this was a realistic volume of data, storage capacities have since rapidly moved on. The result of this would be that the ability of the tools in each case study to handle realistic volumes of data would not be realistically reflected. In order to mitigate this disadvantage, this researcher used a disk image of his own personal laptop. This provided data spanning a year and a half, and contains a volume of around 400GB of data. This dataset would primarily be used to assess the ability of various tools to display a large amount of data in meaningful format without obscuring key points, and to do so in a timely fashion. As this data also contains personal information, it was treated in the same way as any donor images previously mentioned.

All datasets were stored in their raw binary format as is output from common open-source imaging tools such as ‘dd’, a well-known tool used to create bit-by-bit duplicate of a source disk. Although this full copy takes up quite a large amount of storage space, it means that all of the data on the disk is preserved, even if previously deleted by the user. It also simplifies processing of the data, as some other proprietary imaging tools create output formats, which although they may be smaller in size, can only be processed by a limited number of other software tools and have limited documentation. One example is the EWF format used by EnCase Forensics. (ForensicsWiki, 2015).

3.2 Dataset Pre-Processing

In their raw forms, full disk images present the end user with too much incomprehensible and largely unusable data. These raw datasets will not be presented to the end user, as it is not the aim of this research to create a tool which can process raw disk images, as a number of tools which successfully serve this purpose already exist. It is for this reason that it was decided that a tool would be used which was already capable of processing the dataset into usable information, yet lacked the ability to effectively visualise this information.

As was briefly mentioned in the previous chapter, Autopsy 2 is a well-known, open-source digital forensics tool which evolved from “The Sleuth Kit”, a set of command line tools to aid in digital forensic analysis. The first version of Autopsy was a multi-platform, web based tool, which provided a graphical front-end to The Sleuth Kit. However, the newest version of Autopsy, Autopsy 3, is a Windows only tool, and is entirely self-contained.

It was decided that as Autopsy 3 is a mature tool with developed processing features, this would be used for pre-processing the disk images prior to using them as an input to the developed visualisation tool. This is further assisted by the fact that instead of outputting its results into a proprietary format such as many commercial tools do; its output is saved into a well-documented SQLite (SQLite, 2016) database format. This database holds a significant amount of metadata about the results it has found, and also metadata about every file on the filesystem. In turn, this means that this database can be queried and a substantial amount of information retrieved. Should it be necessary to display the actual content of the file, the database holds information about where it is held in the source disk image. This approach is preferential because the database effectively operates as an index for the disk image. As such, writing applications which analyse the data is simplified, as SQL statements can be used

rather than interacting with the raw file-system of a disk image. This may also lead to performance improvements as the database engine will only return data which has specifically been requested. The format is also convenient as it is a self-contained file-based database which means that any application which implements the SQLite library can interact with the database. The distinct advantage of this is that it allows the database to be moved between systems and quickly used without requiring that database engine software first be installed. This is the approach that is taken with more powerful database software such as Microsoft SQL Server, Oracle Database and MySQL. By using this format, it also meant that although the Autopsy 3 application itself was a Windows-only program, the SQLite database produced by it is cross-platform, and as the Autopsy 3 software is only used once to pre-process the disk image, the visualisation software prototype could be written to run on any operating system. Although this may not be an ideal scenario in a production scenario, which creating a proof-of-concept tool it is an acceptable compromise. If this tool eventually came to be worthy of release to the general public, with a little work it could be ported to different operating systems or adapted to interpret different input formats produced by pre-processor tools predominant on other operating systems. I.e. Autopsy 2 which is Linux based.

One of the significant advantages afforded by Autopsy is the number of different ‘ingest’ modules provided by the tool. These modules each looks for different kinds of information in the source disk image. For example, at the time of writing, the modules included could process information such as EXIF Metadata (information attached to image files), hash lookup (matches file hashes to a blacklist), keyword search etc. (Figure 3.1 - Autopsy SQLite database table structure). This functionality can also be extended by third parties who can write ingest modules which plug seamlessly into Autopsy. All of these ingest modules output their results into the SQLite database without altering its structure (Figure 3.2). This is

important when using the tool as a pre-processor; as it means that any additional features which are added to the application will not significantly alter the output format. Thus, a proof of concept tool can be written against this output format, and any new information from the pre-processor can be either ignored, or incorporated into the tool with minimal effort.

Name	
▲	Tables (20)
▷	blackboard_artifact_tags
▷	blackboard_artifact_types
▷	blackboard_artifacts
▷	blackboard_attribute_types
▷	blackboard_attributes
▷	content_tags
▷	reports
▷	tag_names
▷	tsk_db_info
▷	tsk_file_layout
▷	tsk_files
▷	tsk_files_derived
▷	tsk_files_derived_method
▷	tsk_files_path
▷	tsk_fs_info
▷	tsk_image_info
▷	tsk_image_names
▷	tsk_objects
▷	tsk_vs_info
▷	tsk_vs_parts

Figure 3.1 - Autopsy SQLite database table structure

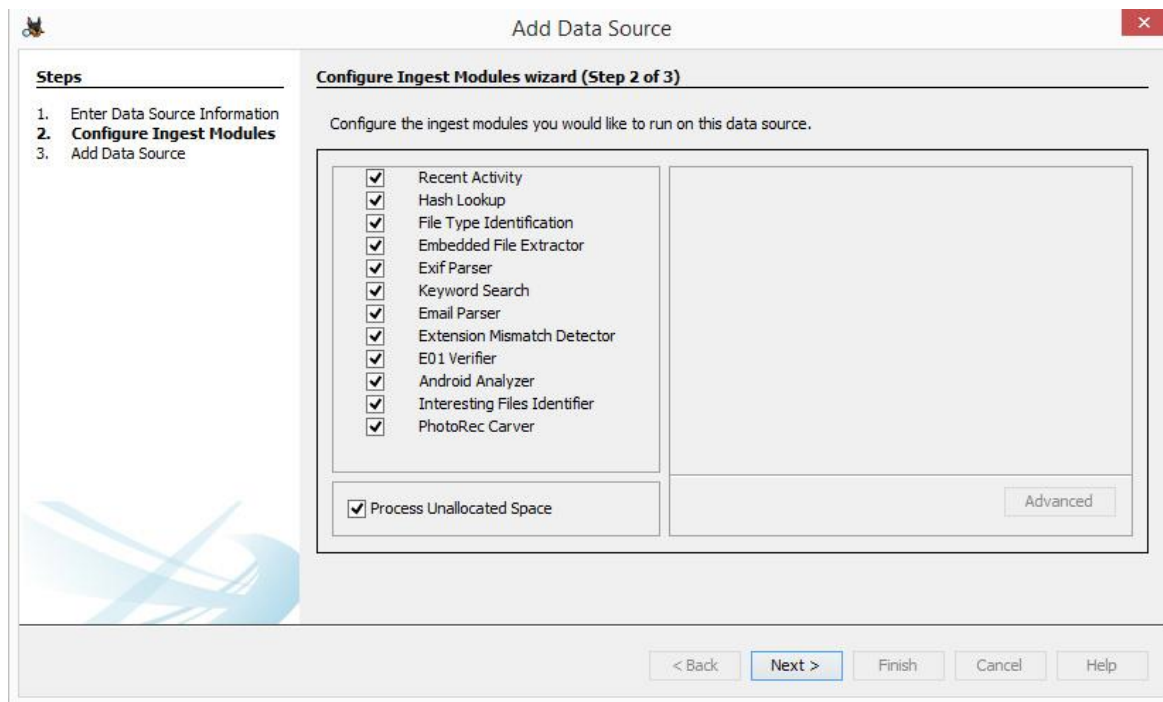


Figure 3.2 - Autopsy Ingest Modules

3.3 Experimental Approach

The purpose of this research was to assess whether any benefits can be realised through utilisation of information visualisation techniques in digital forensic analysis, especially in terms of improvements in investigative effectiveness when examining a suspected criminal dataset. In order to measure the effectiveness of an investigation, this research splits this into two distinct metrics. The first of these metrics is the efficiency of the investigation which is measured by how long the investigation takes to complete. The second metric is the accuracy of the investigation. This measures how often an investigators conclusions match the actual truth behind a dataset.

The methodology which was be used in this research was that of the use of a number of case studies. Each case study examined the use of a built prototype, or a pre-existing tool. The outcomes of each case study look at how the digital forensics information is displayed to the end user, the benefits which can be taken from the case study, and any areas which require improvement or which render the tool unsuitable for application in a digital forensics context. This assessment took the merits of the tools and looked at whether they were a useful tool in their own right, with a view to creating a prototype visualisation tool which took the lessons learned from each case study and applied them in a way which would be useful to the end user. This final stage prototype tool would then be tested with a group of users knowledgeable in the area of digital forensics.

When looking at each case study, it was important to look at how each performed in various scenarios and at the ways in which each presented the information to the end user. With information visualisation, it can sometimes be the case that a visual model looks appealing

and unique, but when subjected to further analysis and user testing, it fails to present the information in a way that is comprehensible, useful, or complete. This is a scenarios which must be avoided, as unclear or missing information in a digital forensics investigation can lead to erroneous or incomplete conclusions being drawn. In this situation, if these conclusions are presented during a criminal court case, it could have a substantial impact on the outcome of the case.

This research aimed to look specifically at the exploratory value of each tool; that is, not whether the tool can present a set of pre-formulated conclusions to the end user, or whether it can automatically validate the end users assumption. Instead it was to be assessed if the tool gives the end user the ability to freely explore the data in a meaningful way, and to derive their own conclusions. An exploratory approach (Liu et al, 2013) to information visualisation opens new opportunities for the user to discover facets of the information they were not previously aware of, and to derive patterns from disjoint fragments of information which can be hidden by tools which present the information to the user in a static and pre-formatted fashion.

In the first instance, each case study was examined using the ‘John Doe’ test case which was previously mentioned. This dataset contains a relatively small amount of information; providing a set of events which occur across a 2 week timeframe. This dataset was used so that the tool being examined was not overly stressed in terms of processing time and resources required. Instead, it allows the known narratives of behaviour to be viewed in the context of the tool, to examine whether all necessary information is shown to the user, and whether these narratives are visible when using the tool. The John Doe dataset was also used

to test the final prototype tool with a test group, as it gives base truth which we could test against, thus providing a controlled environment. The scale of the data also provides the ability to test the tool with a user group relatively quickly. Using a larger dataset would not be feasible as, not only would it be difficult and time-consuming to distribute a large dataset to the test users, but it would also take substantially longer for the users to work through the dataset. This would effectively preclude the ability for the users to conduct any experiments in a single sitting; a necessary attribute as tests would need to be conducted in a controlled environment, with the assurance that participants did not have the opportunity to discuss the experiment at length with each other. Although realistically it would be unlikely that an investigator would work entirely in isolation, for the purposes of this research, it is necessary in order for us to gain a measure with which we can validate the hypothesis against.

The ‘John Doe’ dataset, although a useful dataset to work with when examining digital forensics tools, does not give a realistic impression of how a tool will perform when asked to process and represent a dataset of a modern size; as would often be encountered by digital forensics investigators. The other dataset used for this was a raw image of a 500GB hard drive which had been frequently used for around 2 years. As such, there was a large volume of events, and a large number of deleted files which could also be examined. The sheer volume of this dataset, and the fact that it had not been used for any criminal activity as such meant that it was difficult to draw any specific narratives of behaviour from the dataset which would be meaningful in user tests. Instead, it was effectively used as a ‘load-testing’ dataset. Each case tool was given this dataset, and its performance was assessed. This was not studied to the point of strictly timing the tools and creating statistical data from these timings. Instead, the approach was taken of merely checking if the tool would load the dataset within what would be considered to be an acceptable time, and if it continued to operate without a

problem while the dataset was browsed by the user. It was also important for this research to examine how the visualisation would perform when given such a large database. Specifically, to ensure that the information was still clear and could be effectively analysed by the end user. The results of studies using this dataset shall be discussed, however, due to privacy concerns, the dataset itself is not available to the public and any extracts from the dataset will be presented in a redacted format. It is not feasible to release this entire dataset in a redacted format, as due to its volume, it would be an almost impossible task to ensure that all personally identifiable information was censored. This was not an issue with the John Doe test case, as this dataset had been created on a machine which had been forensically erased and then used only for the sole purpose of creating the dataset.

Prior to any final experiments being conducted, a small informal pilot study would be conducted with a number of suitably qualified volunteers. This would ask the participants to find a few small pieces of information from the dataset, with some guidance as to the operation of the application being provided. The main purpose of this being to provide a form of user acceptance testing. This pilot test would allow for small bugs and design flaws to be corrected before a full-scale user experiment was undertaken. The importance of this is that any bugs which remained when conducting a formal experiment may significantly skew the results, with users placing focus on errors in the application as opposed to the format of the tool itself.

The final experiment as part of this research involved the use of a suitably trained participant group who were presented with a number of tasks to complete which required them to analyse the John Doe dataset; with either the Autopsy tool or Insight tool, depending on

which they had been allocated. These tasks focused on different areas of the dataset, and the participants would be asked to find the correct answers from the entire dataset. At the end of each task, the participant would be asked to rate, on a scale of 1 to 5, how easy they found the task to complete, and asked for some feedback into why they felt this way. The participants would also be timed to determine how long it took each of them to complete the experiment. The time to complete the experiment would be used as the efficiency metric for the purpose of this research, to determine whether the visual nature of the prototype tool allowed participants to complete their investigation more rapidly. The correctness of the participant's answers to the tasks would serve as the accuracy metric, allowing conclusions to be drawn as to whether the participants were finding the correct information in the dataset. These two metrics would be the primary way to validate the research hypothesis, with an improvement in either metric being interpreted as an improvement in effectiveness. However, the feedback in the form of the Likert scale data, and the feedback serves as a way to gain further insight into the results, showing if there were specific reasons for data patterns. For example, if there is a drastic shift in accuracy for one task, is this because the participants found the tool especially difficult to use?

3.4 Conclusion

The methodology of this research takes the form of number of pilot studies conducted by the researcher, which would examine case studies of current pre-existing tools and prototypes designed by the researcher. These case studies would provide outcomes which would lead the direction of the next case study, in combination with ideas taken from the review of literature. After a number of case studies, the outcomes would be used to design a prototype tool called Insight. This tool would then be subjected to informal testing with 5 Ethical Hacking students to assess the functionality of the application, gain preliminary feedback, and to identify any oversights which affected usability. The application was then altered to correct these oversights before being subjected to a full-scale experiment with 29 participants. The results from this experiment would then allow the hypothesis to be proved or disproved.

The outline of this methodology can be described as:

1. Select format of first dataset based on literature review.
2. Researcher to assess how clearly the case study tool displays user to the end user.
3. Outcomes of case study summarised.
4. Outcomes used to decide format of next case study.
5. Cycle repeated until a significant number of outcomes are drawn which allow the development of a custom prototype tool. These outcomes will dictate the format of the tool, and way in which it presents the dataset to the end user.
6. Development of prototype tool.
7. Informal user acceptance testing.
8. Refinement of prototype tool based on informal user tests.
9. Full-scale user experiment involving suitably trained participants.
10. Analysis of experiment based on efficiency and accuracy metrics, and also thematic analysis of user feedback.

Chapter 4 – Case Studies

4.1 Introduction

This chapter will discuss the case studies which looked to examine a number of visualisation techniques, and their application to a digital forensics dataset. The lessons learned from each case study in regards to the clarity of the information when displayed to the end user, and how useful it would be to an end user will be discussed. This will include an analysis of the strengths and weaknesses of each method, which will be fed into the next study, and provide a framework for the development of the final visualisation tool which will be used to draw results for validation of the research hypothesis.

4.2 Case Study 1 - Timecubes - Pysight

4.2.1 Introduction

This case study aimed to review the available visualisation formats, and specifically, how the well-known timeline format could be adapted to be displayed in 3D space. In doing so, it was envisioned that there would be more space to display events, and would be easier for the end user to view different types of events occurring in close proximity to each other.

4.2.2 Development

After reviewing the literature, and finding that the overwhelming majority had focused solely on 2D techniques, it was decided that it would be worthwhile to pursue a tool which would represent the dataset information in 3D space, as this approach had little previous research. It was envisioned that by utilising 3D methods, more space would be available to display the dataset, which could make analysis more intuitive for the end user. As such, it was

investigated how it would be possible to create a basic 3D model, relatively quickly, so that it could be assessed whether there were any benefits likely to be granted by using a 3D model.

At first, the Java 3D (Selman, 2002) framework was examined to test whether it would be a suitable platform for development. Although the Java language itself is still heavily used and well supported, the Java 3D libraries were less well so. The latest stable release of these extensions were around 7 years old at the time of writing. This was reflected in its relatively complex setup and verbose coding style. In the search for an alternative development platform, it was decided that Python would be the ideal choice for this. There were many reasons for this decision, one of the main being that it is a language which is designed for rapid development of scripts, with its large standard library, and simple coding style. It was also found to be very easy to import code written by others. The Python language, however, does not natively provide the end user with libraries suited for the development of 3D graphics development. For this reason, the VPython library was used. This allows for basic 3D applications to be developed very quickly. Where Java3D gave flexibility through granular code, requiring each part of the scene declared; VPython allowed simple code to quickly generate a scene. For example, simply using the code `'sphere()'` would create a grey sphere which the user could rotate around. Clearly this is of limited use on its own; however, by applying various attributes such as size, position and colour to these primitive shapes, it opened up the possibility of a rapid visualisation design methodology, where various models could be created, modified and tested in a much shorter space of time than would usually be possible with traditional frameworks.

One of the areas that is often of high value to digital forensics investigators is the web browsing history stored by the browser software on the device. This is evidenced by a multitude of various tools existing which specifically target this area. For example, Web Historian (Lowman & Ferguson, 2011), Webscavator etc. It therefore made sense to try and create a basic tool which would show this data as a starting point. It was envisioned that with the additional 'space' given through the use of a 3D format, it would be possible to add other data sources to the model. A "timecube" format was decided to be the style of this visualisation, and a proof of concept tool would be built to test such a model.

4.2.3 The 'Timecube' model

The "timecube" model was envisioned to be a novel visualisation style which would take the well-known and significantly well-established style of the standard timeline model, and adapt it to make use of 3D space. The application would feature a flat 2D panel, which would depict web browsing events as dots. Each dot would be placed on the X axis based on the date it occurred, and would be placed on the Y axis, based on the time it occurred. Through this format, the user could easily read up the Y axis to look for patterns occurring on a certain date, or along the X axis based on a time. This would make it simple of the user to spot if there were large numbers of web browsing events around certain times or dates. This mirrored the approach taken by the research by Lowman and Ferguson (2011) in their Webscavator tool (Figure 4.1).

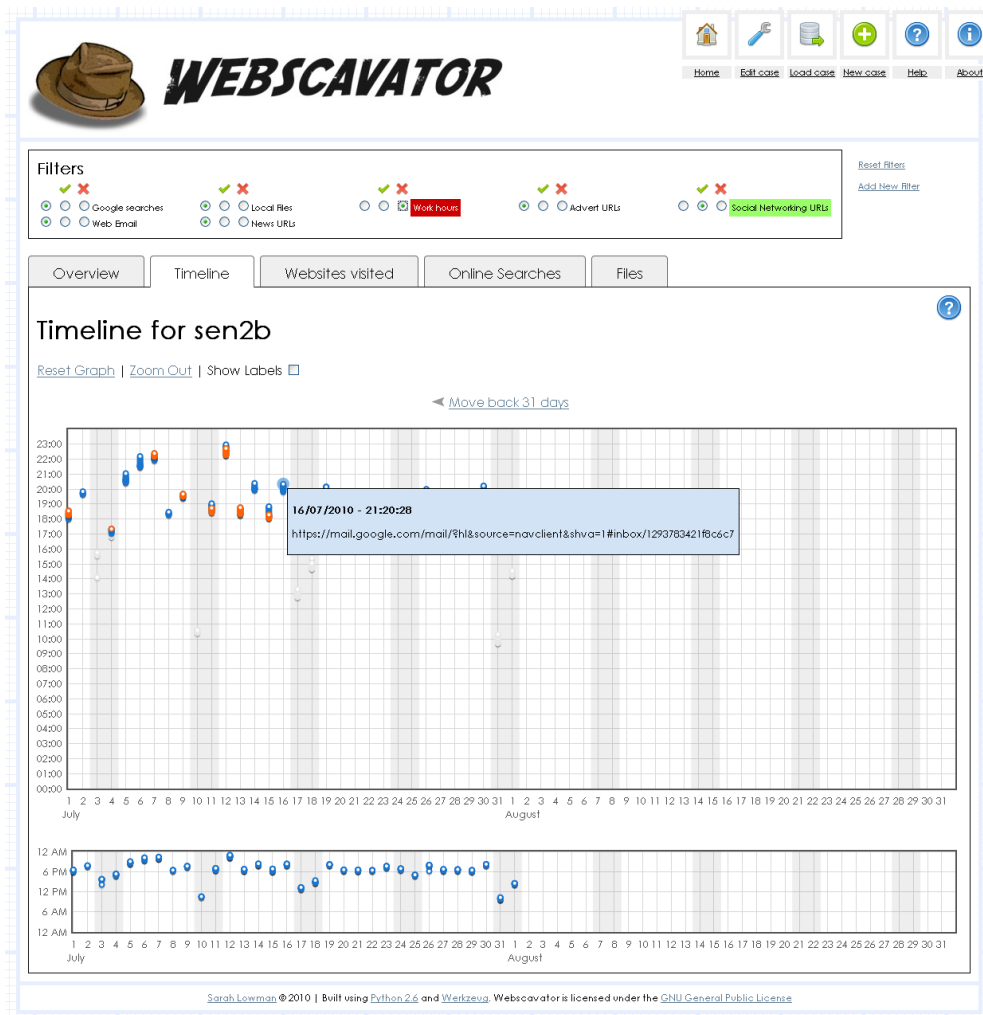


Figure 4.1 - Webscavator Timeline View (webscavator.org)

This was then extended by following the same approach for other temporal data points from the dataset. Some examples of these data points include key file creation or modification times, Windows event log data etc. These other data sources were placed on a 2D plane of their own, sitting underneath the plane holding the web browsing information, Each layer was made partially transparent to allow the other types to be seen (Figure 4.2), and each dataset was represented with different coloured points so they could be distinguished from the other datasets quickly. The theory behind this approach was that, by allowing the data points to overlap without them being completely obscured, the end user may be able to see patterns and similarities emerging between them.

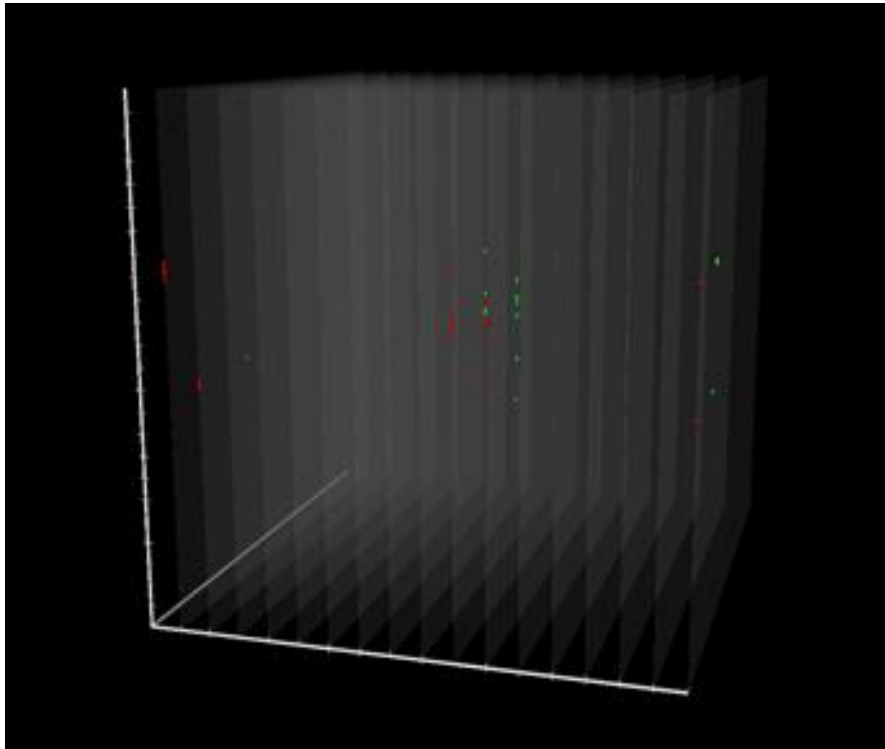


Figure 4.2 - Timecube 'layers' representing days

4.2.4 Prototype Testing and Evaluation

The Pysight prototype was tested by means of a pilot study by the researcher, primarily with the John Doe dataset, adding progressively more data to the visualisation in order to test how clear the information was to the end user. The first dataset to be added to the model was the web browsing history of the device as this was the type of data shown to be successfully visualised using Webscavator, and also generally provides a relatively clear narrative of user behaviour. Visualising this information a single plane was quite successful, as it was quick to see dates and times of when the user had been accessing the internet. This however, provided no new results to us, as this effectively was what had been shown in previous research as mentioned. With only one dataset shown, there was effectively no use of the 3D plane on the Z axis.

The next step in the research was to add another small dataset, such as JPEG creation times to the model; with a higher Z value, meaning the data points would initially appear behind the existing web browsing history points. By changing the alpha channel value of the first dataset, it was made partially transparent so that the second dataset could be seen. This was useful as the investigator was able to see if the device user had perhaps downloaded any images while browsing the internet. A third set of data points were then added to the visualisation. It was at this point that it started to become difficult to distinguish individual data points, as the transparency values of each data point started to add up and make the data points which were the furthest away from the user, more opaque. As such, data points which were represented by colours which were not as vivid, or appeared to be more transparent to the user, were difficult to distinguish when in a densely populated area of the 'cube'.



Figure 4.3 – Side view of the time cube

Attempts were then made to determine if by allowing the user to freely rotate and zoom the cube, these data points would become clearer. In examining this, it was found that when the end user was looking directly at the “front” of the cube as it was originally rendered; patterns of behaviour were visible on the first few layers. When the cube was then rotated and zoomed, the user could see through the spaces added between the data layers, making all data points more visible. In doing so, however, the user lost almost all ability to read the chronological information of the model. It became extremely difficult to see on what days and times certain events occurred. Arguably this information could have been restored to the user through the “details-on-demand” approach in which the user would click a data point to see the date and time of which it occurred, however, this is generally not an efficient way to

present this information, as this approach is generally reserved for auxiliary information which would be beneficial to the user, but which does not affect the overall context of the information. This problem was further amplified by the fact that when the user rotated the cube, the user's perception of patterns between the points changed dramatically. In looking at the information from a different angle, data points appeared to line up in different way, creating patterns which were meaningless, and making meaningful patterns invisible to the user. Combined with the almost complete loss of chronological context, this made the model very unintuitive, and risked the user losing their sense of "which way was up".

From a performance perspective, this visual model was quite slow to load. This was not necessarily a fault of the libraries or language used, but in fact the approach to how the data was accessed at runtime. This proof of concept application was developed with the view that the database could change, and as a result, at each launch the application would query the SQLite database for the information it needed. This took a minute or so for the small John Doe test case, however, when the larger test dataset was loaded, the application became very inconvenient to use due the extremely long loading times, reaching into the 15 min range or longer, depending on how many types of data were being loaded. Long loading times could increase investigation times, and makes running experiments with the tool difficult. Should the tool become unresponsive, participants of the experiment would be required to wait until the tool has loaded again before they could resume.

4.2.5 Case Study Outcomes

The primary point taken from this study was the effect that visual perspective plays on how the data points are interpreted by the end user. By allowing the user to rotate a 3D model, it

has been found that it is important that the user be able to maintain a fixed perception of context in order to ensure that the data points retain their meaning. For example, if the user rotates the entire model 180 degrees, it should be obvious to them that they have done so. Failing to do this may substantially alter the understanding of the data; as was the case in this study, whereby such an alteration entirely flips the chronological scale. Regardless of this, it is also a challenge to ensure that the user does not rotate the model into a position in which misleading patterns are displayed; a situation which could lead to erroneous conclusions being drawn, and factual conclusions being entirely missed.

The performance aspect of the visualisation is something which is a little more straightforward to address. From this case study it was realised that in fact, the database is highly unlikely to change as the source data must be protected from modification for it to be forensically sound. For this reason alone, it is entirely unnecessary for the full database to be queried every time the application is launched. This led to some thought into how this could be avoided, and as such, the idea of a separate index of relevant information was explored. This idea would be taken forward in subsequent case studies, with some research into the best format for this index. Ideally this index would be portable between tools, allowing for the continued rapid development of proof of concept case studies.

4.3 Case Study 2 - Hyperbolic graphs - Walrus

4.3.1 Introduction

For the next case study, it had been considered whether it would be useful to focus on filesystem data, and if this could be represented in way onto which other information could be overlaid, for example, file creation times. 2D radial trees were considered as an option, as they seemed a natural evolution of the traditional file system hierarchy model. However, it was decided after some research that they would be likely to suffer similar downfalls, such as cluttering when there are many data points at the same level of the hierarchy. As an alternative, it was decided that a hyperbolic visualisation approach, as demonstrated by Munzer (1995) and discussed in the Literature Review, would be examined. As this tool would not use chronological information to determine the data points, it should not suffer from the same skewing of behavioural patterns and loss of context which became apparent during the Pysight case study. The space provided using 3D methods would provide more space to display the data.

The value of such a format of tool may be that it could allow forensics investigators to examine two device filesystems side by side, and quickly spot any differences or patterns between the two. This feature may be useful in any scenario in which the end user needs to quickly check if there have been any significant alterations to a system between two points. One scenario in which this could be especially useful is if a system is compromised by an external intruder. Using this type of visualisation, it could potentially make it simple to spot if the intruder created, modified, or removed any part of the filesystem; regardless of whether the changes were made close to the root of the hierarchy or many levels deep.

In order to test hyperbolic visualisation methods, a tool, developed by Young Hyun (Hughes, Hyun & Liberles, 2004) at CAIDA, called Walrus was used. This is an open-source tool, based on hyperbolic visualisation research conducted by Tamara Munzner. The tool generates hyperbolic directed graphs based on an input file in their LibSea format, with the ability to attach labels to the nodes and to define their colour. The use of this tool allows us to perform a case study into how hyperbolic visualisation techniques could be applied to digital forensics datasets, while removing the need for a complex tool incorporating these techniques to be developed from the ground up. The LibSea file format is similar to XML or JSON, in that it is human readable to the extent that it can be opened in a text editor and modified simply (Figure 4.4). It defines basic information about the graph such as the links between nodes, and various attributes such as the labels and colours of the nodes. This file is then automatically used by Walrus to generate a hyperbolic model (Figure 4.5).

```
Graph
{
  @name="Directory Tree";
  @description="A graph generated from a directory tree.";
  @numNodes=1450;
  @numLinks=1449;
  @numPaths=0;
  @numPathLinks=0;

  @links=[
    {@source=0; @destination=1;},
    {@source=0; @destination=2;},
    {@source=0; @destination=3;},
    {@source=3; @destination=4;},
    {@source=4; @destination=5;},
    {@source=5; @destination=6;},
    {@source=6; @destination=7;},
    {@source=7; @destination=8;},
    {@source=8; @destination=9;},
    {@source=9; @destination=10;},
    {@source=5; @destination=11;},
    {@source=11; @destination=12;},
    {@source=12; @destination=13;},
    {@source=13; @destination=14;},
    {@source=12; @destination=15;},
    {@source=15; @destination=16;},
    {@source=15; @destination=17;},
    {@source=11; @destination=18;},
    {@source=11; @destination=19;},
    {@source=11; @destination=20;},
    {@source=20; @destination=21;},
```

Figure 4.4 - Snippet of LibSea graph file format

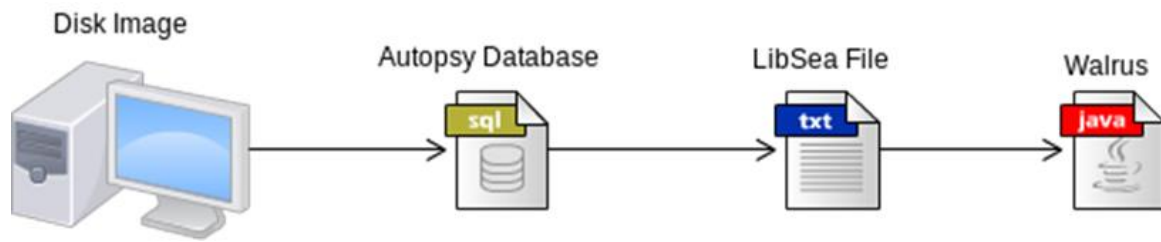


Figure 4.5 - Disk Image to Walrus Stages

In order to generate this file a Python script was written to extract information about the filesystem from the Autopsy database, and to write this out into LibSea format. The format of the Autopsy database made this a simple task, as each element of the filesystem is given a unique numerical identifier. As such, it was the task of the Python script to take these identifiers, join them to their parent identifiers, and to output them into the LibSea format. Although this was relatively straightforward, the process itself was time-consuming due to the fact that the script had to recursively query the database. When the script was tested on the 450MB Autopsy database generated by the large 500GB dataset, it took over 15 hours to generate a LibSea file. However, this figure should be taken lightly, as the code was in a prototype stage and had not been written with the most efficient operation in mind. It was hypothesised that a large amount of this time was due to disk operations. To test this, the entire Autopsy database was copied to an in-memory database. This meant that the entire database would only be read from a slow hard disk once, with any subsequent operations taking place on high-speed RAM. This had the result of reducing the time taken to process this specific dataset by 5 hours. Undoubtedly, further optimisation would have reduced this figure further. The LibSea file generated from this dataset totalled 313,000 lines.

After a variety of disk images were processed into LibSea format, they were loaded into Walrus in order to determine whether they displayed any information which may be useful to the end user; specifically whether it was possible to identify directory structures and other facts, simply by looking at the hyperbolic model. As part of this case study, 2 disk images were taken from devices which had clean installs of Windows. One of these images was of a Windows XP device, and the other a Windows 7 device. When loaded into Walrus, the difference between the two models was instantly noticeable (Figure 4.6).

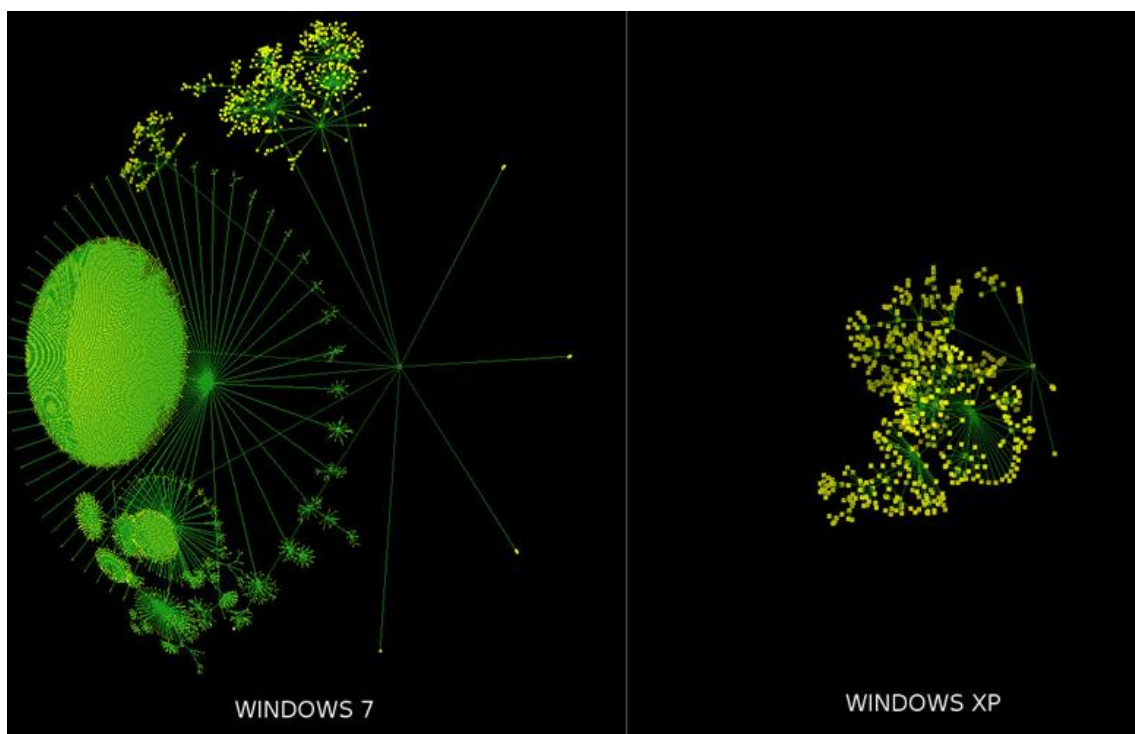


Figure 4.6 - Windows 7 and XP clean installs in Walrus

The experience of this researcher is that when given the ability to select nodes and see their directory path, it is possible to form an ability to recognise directory structure based on the patterns generated by the hyperbolic model. For example, from the Windows 7 model shown on the left hand side of Figure 4.6 we can see a very large cluster of directories, all branching

from one parent directory. Having examined this structure using Walrus previously, it is quick to recognise this as the WinSxS directory. This directory is the component store for Windows, which holds a large number of core system files and continues to grow over time as system updates and other applications are installed. At a basic level, the size and presence of this folder can give us indications at a glance of which operating system is installed.

However, a potentially more useful feature available to us in Walrus is the ability to define the colour of individual nodes in the model (Figure 4.7). A pilot study was conducted by the researcher with this feature set to represent timestamp information. The Autopsy database was queried for the timestamp of the most recently created file in each directory. This timestamp was then incorporated into the LibSea graph file as an attribute for each respective node. When the graph is generated, each node is represented with a colour based on the attribute value defined in the LibSea file. Through this approach, there is a potential for the end user to recognise directories which are frequently used by the end user, or in the case of the WinSxS structure mentioned previously, whether any system updates or service packs have recently been installed. It is hypothesised that this information could rapidly lead the investigator to areas of the directory structure where the device user has been frequently using or modifying files, perhaps hidden deep down the directory tree as an attempt to evade detection. Although in this case study, the timestamp itself was used as the colour value, theoretically when the timestamp is being processed by the script to generate the LibSea format, a conditional statement could be used to define the colour to use. For example, files which were modified within the last month could be set as red, with older files being set as yellow or green depending on the time frames selected. Another interesting application could be to have the Python script compare the filesystem from a device at two different points in time. Colours could then be used to represent changes between the two points in time, i.e.

whether a directory has been deleted, added or significantly modified. This could draw inspiration from the Change-Link software developed by Such an approach would allow an investigator to focus their efforts on interesting areas of the filesystem, or on devices which show frequently changing content. This would be a particularly useful method for law-enforcement agents to target their searches into areas which are frequently used by the device owner.

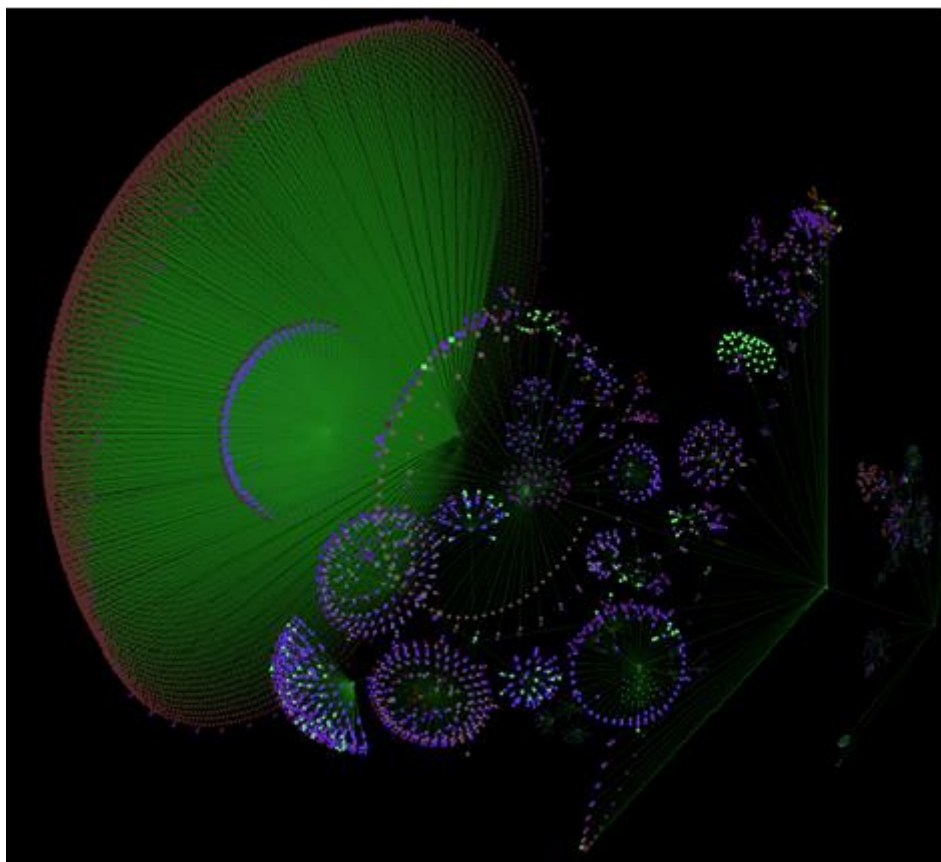


Figure 4.7 - Walrus nodes coloured by file creation timestamp

4.3.2 Prototype Testing and Evaluation

The first point to note regarding this visual model is that, unlike the Pysight case study prototype tool, this tool does not provide the end user with any view of event type data. This

includes the web browser history, system events etc. These events do not lend themselves well to a model which represents hierarchical data structures. Instead, the focus of testing was on recognition of common file system structures based on their location the hyperbolic model, and on their colour. This approach was taken to assess whether this different type of information could be effectively visualised in a 3D space, and whether it could be as useful to an investigator as event-based data.

As the previous case study suffered greatly from occlusion of information, this was the first test to be conducted. The graph can be freely rotated and centred on nodes of the user's choice. These features, combined with the fact that the hyperbolic model layout adapts to accommodate changes such as different nodes being selected as the root, and as a result all of the nodes stay visible to the user. This is especially relevant when using a graph which has densely populated areas, such as was discussed previously regarding the WinSxS folder structure. If the user wishes to see the contents of this folder structure more clearly, they are free to select the WinSxS folder as the new root node of the graph. Walrus will then relocate all of the other nodes and space them out appropriately, allowing the individual subdirectories contained within the WinSxS directory to be distinguished more easily. The disadvantage of this approach, is that there is the potential for the patterns of directory structures previously recognised by the end user to be changed to a point where they are more difficult to recognise. However, in testing, it was apparent that these patterns are still recognisable in most cases, albeit slightly smaller and more difficult to see. Larger structures such as the WinSxS structure are easy to see because of their dense cluster of nodes. When the nodes are coloured, this also assists in the ability to recognise previously encountered directory patterns, as in cases where the graph has been altered to a point where the links between

nodes is not as clear, the colour patterns of the nodes remains clear, providing the user with a sense of context.

The second consideration which was brought forward from the PySight case study was that the user's viewpoint of the model can substantially change their perception of patterns within the dataset. When using Walrus, this can still be an issue to an extent, as the user is free to rotate the model as they please, and as such it may not be as simple to spot certain previously recognised patterns if their location or rotation has changed substantially. However, this issue is mitigated to an extent by the fact that all nodes in the hyperbolic graph are visually linked together, thus maintaining the user's perception of the relationship between the nodes. This was an issue for the previous case study, as the events had no explicit visual links, relying entirely on their positioning to reflect this. Furthermore, the ability to maintain pattern recognition is further assisted when colour is used; whereby, regardless of viewpoint, the user retains the ability to recognise patterns of colours they may have observed before.

From a performance perspective, the pre-processing stage of this application took far longer than would be in a production tool, however, as a case study, this should not be counted against the tool. Empirical evidence shows that there is most certainly a chance that this pre-processing stage could be drastically reduced. Simply loading the Autopsy database entirely into RAM through the use of SQLite's In-Memory Database feature led to an immediate reduction of the processing time by a third, as was discussed previously. Further optimisations to the SQL queries performed by the application would be likely to afford even greater performance gains.

4.3.3 Case Study Outcomes

From the previous case study, we could see that as the application needed to retrieve information from the Autopsy database each time the application was loaded, thus leading to degraded performance as a result. From this outcome, we drew the conclusion that a form of index would be useful so that the application could quickly load, and then retrieve any additional information as and when the end user required it (the ‘details-on-demand’ part of the visual information-seeking mantra of Schneiderman (1996)). The Walrus application by its very operation requires this ‘index’, accepting only a graph file that contains the numerical identifiers of each piece of filesystem data, and the links between them. The structure of this format means that when declaring the relationship between the nodes, auxiliary information such as the actual directory name is not provided. This prevents repetition as multiple links are declared. Instead, this information is held at the end of the file, after all of the links have been declared. Such an approach allows the application to draw the hyperbolic graph structure, with no real knowledge of what the graph represents, and then allows it to fill additional information such as colour and labels after the graph itself has been constructed. Although it was time-consuming for a script to generate this graph file; it was a one-time process. Once the file had been generated, it could be loaded into the Walrus application and a visual model displayed to the end user in less than a minute. It could be argued that this approach is inflexible, as it does not allow the dataset to change without requiring the entire index file be regenerated. However, the counter-argument to this would be that digital forensics by its very nature, is about the analysis of a snapshot of a device; a frozen copy of the data contained on it, which if changed would no longer be admissible as evidence in an investigation..

As this visualisation focused on the filesystem of the device instead of the event data shown by the previous case study, it was of interest of this research to conduct a brief analysis of whether this type of data would be of value to an investigator. It is relatively unlikely that filesystem metadata data alone would be of great use in a digital forensics investigation being conducted for the purposes of law enforcement. The reasoning behind this hypothesis is that the narratives of behaviour behind the data, and the motives which led to the state of the filesystem which is being examined, are of much greater interest to a law-enforcement agent. Such information allows them to prove in court that the motivation to commit a crime was present, that the suspect knowingly did so, and that there are no mitigating circumstances surrounding the evidence presented. Such information is almost entirely provided by event log type data rather than the filesystem data alone. Filesystem information may be of use to a in the case of digital forensics in a corporate setting, however, this research does not seek to address this field.

The main industry of interest for this research is that of law enforcement. The reasoning behind this is, as can be seen from the literature review, the law enforcement community in general is struggling to keep pace with the increasing amount of digital evidence they are being required to review. As such, although this hyperbolic visualisation of filesystem data may be useful in some instances if developed further; as an outcome of this case study, we conclude that in its current state as a standalone tool, it would not further the interests of the law enforcement community. However, what can be taken from this case study is the LibSea file format; specifically, the fact that it is a file which holds an index of the data, which can be used to quickly generate a visual model. Although this is a specific format for the Walrus tool, which is designed to store hierarchical datasets, we would take this as an inspiration for moving forward with the subsequent case studies.

4.4 Case Study 3 - Timeline - Insight.js

4.4.1 Introduction

When considering an approach for the next case study, it was decided that the prototype tool would be developed as a web application rather than a desktop application as in the previous case studies. The reasoning for this was that in the previous case studies, the application itself often required quite a significant amount of setup before it could be run. For example, the Pysight application required that Python be installed on the system, and that the VPython module be installed. The same applies to the Walrus software which required that the Java Development Kit and Java3D software be installed on the system. The benefit of developing a web application would be that the end user would be able to use it on any computer without requiring that any additional software be installed. It could also have the benefit of being served from a central server. In a law enforcement scenario, this could mean that multiple investigators could all work on the same case simultaneously without each requiring their own copy of the dataset. Software updates would also only require that the new version of the web application be installed on the server, ensuring all investigators had access to the same information and features.

The visualisation in this case study would take the format of a 2D timeline. In this instance, it was decided that the route of a 3D model would not be pursued due to the difficulties faced in pattern recognition when the viewpoint changes. This decision would allow us to explore whether a 2D model would still be able to provide the end user with a way to intuitively explore the dataset, or whether the dataset was too large to be shown on a 2D plane.

4.4.2 Development

As there are a number of JavaScript libraries available which provide visualisation feature sets, it was decided that one of the most popular, D3.js (Bostock, 2016), would be used to create the prototype. The D3.js library is very flexible, in that it provides a framework of features to allow developers to create a varying range of visualisations. The downside to this flexibility, however, is that the code required to create a basic visualisation can be quite verbose, and the style of the code has a relatively high learning curve. As such, the D3.js library quickly became unsuitable for the rapid development of a prototype tool. Instead an alternative library, TimeGlider.js was found. This library provides a comprehensive framework for a timeline visualisation, which can read event data from a JSON file format and represent these events as points on the timeline (Figure 4.8). The timeline which is generated is fully interactive, and allows the user to zoom in to a certain area and to click events to view more detail (Figure 4.9). Each event can also be categorised and represented with a different shape and colour on the timeline. The timeline also provides automatic filtering based on these categories, allowing the user to click on the legend to turn off the display of a certain category of event.

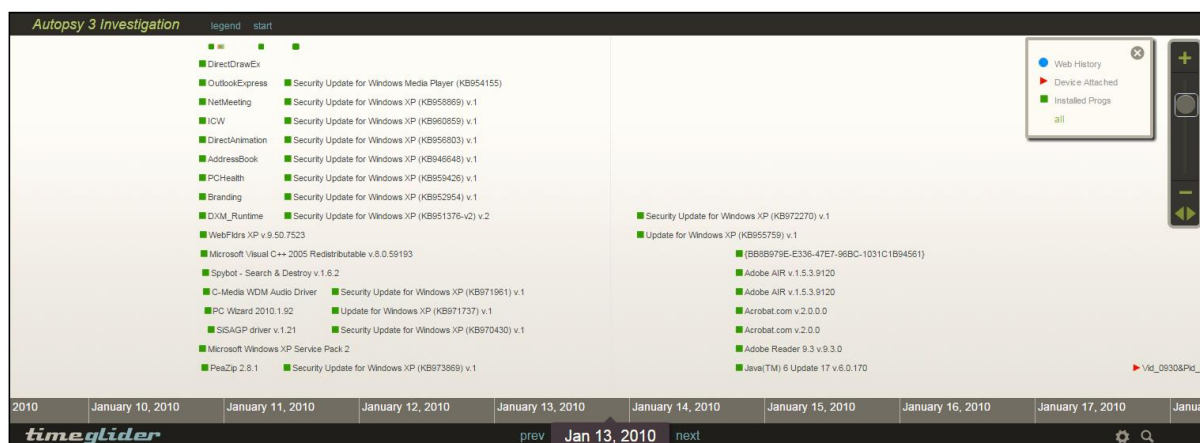


Figure 4.8 – Insight.js Web Timeline

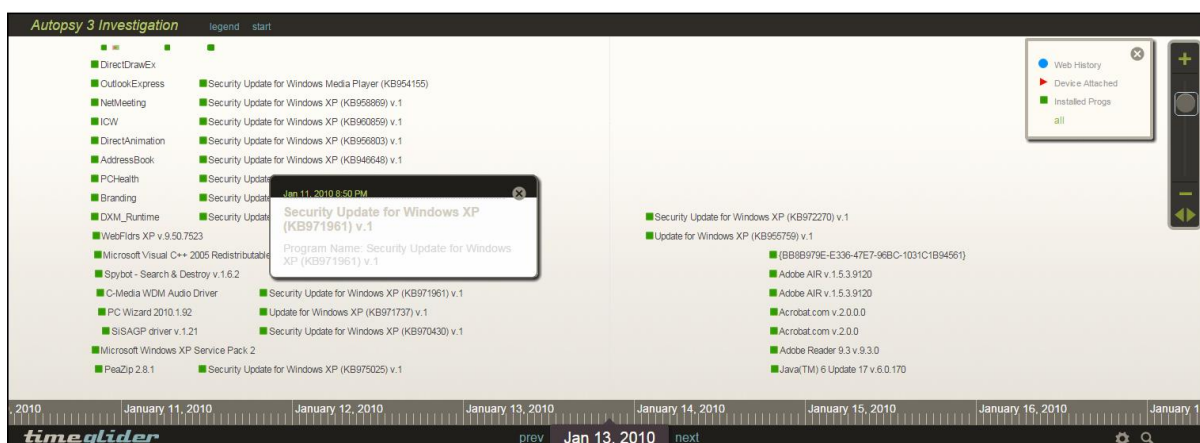


Figure 4.9 – Expanded event detail

As the format accepted by this library was JSON (JavaScript Object Notation), it was necessary to alter the Python script to generate a file in this format instead of LibSea. This also involved changing the kind of information that was being retrieved from the database, as this visualisation was to show system event information, rather than the filesystem information which was being shown in the previous case study. The JSON file held basic information about the event, including a unique identifier which could be used to link it back to the Autopsy database. Each element also had a number of attributes which held

information such as the event timestamp, title, category etc. The category of each event was dictated by the icon chosen to represent it, with user-friendly labels for these categories defined at the end of the JSON file. A small section of the JSON file, including the legend definition can be seen in Figure 4.10.

```
{
  "id": "021369",
  "title": "ReaAtTopOfMacKinnonPass0930.jpg",
  "description": "Parent Path: /",
  "startdate": "2005-02-09 17:05:44",
  "enddate": "2005-02-09 17:05:44",
  "date_display": "hour",
  "icon": "triangle_blue.png",
  "importance": "2",
  "event_type": "event"
},
{
  "id": "021370",
  "title": "ReaEatingRentalCar.jpg",
  "description": "Parent Path: /",
  "startdate": "2005-02-09 17:05:44",
  "enddate": "2005-02-09 17:05:44",
  "date_display": "hour",
  "icon": "triangle_blue.png",
  "importance": "2",
  "event_type": "event"
},
{
  "id": "021371",
  "title": "ReaRetrievingBakedBeanCanFromTern.jpg",
  "description": "Parent Path: /",
  "startdate": "2005-02-09 17:05:44",
  "enddate": "2005-02-09 17:05:44",
  "date_display": "hour",
  "icon": "triangle_blue.png",
  "importance": "2",
  "event_type": "event"
}
},
"legend": [{"title": "Web History", "icon": "circle_blue.png"}, {"title": "Device Attached", "icon": "triangle_red.png"}, {"title": "Installed Progs", "icon": "square_green.png"}]
```

Figure 4.10 - TimeGlider JSON Format

As the procedure for adding different types of events is quite similar; it was logical that the script to extract the information from the Autopsy database be written in a modular fashion. As an early prototype, the Python script contained a large amount of redundant code, and in fact was one long script. When making modifications to one type of event, it became easy to accidentally modify the wrong event type. Instead, each part of the script which processed a certain event type was broken out into its own method. This meant it was simple to change the order in which events were processed, and quick to find the section of code which dealt with certain event types.

4.4.2 Prototype Testing

As was previously mentioned, this case study differed from the previous two, in that, it looked at a digital forensic visualisation which utilised a 2D model, rather than the 3D methods which had previously been examined. The purpose of this was to determine whether the information was clearer to the end user when using 2D instead of 3D. Also, it differed from the last case study in that it no longer dealt with filesystem information, and instead returned to displaying the same event information that was shown to the user in the first case study.

When this prototype tool was tested, it became apparent that as the user had a limited ability to alter the placement of the events on the visualisation, other than zooming or scrolling, this visualisation would not suffer from the same problems as the previous case studies suffered from. That is, the user would not be able to change the viewpoint to a point where the information shown could be misinterpreted. As the information is displayed to the user on a flat plane, with a clear context provided by the time markers at the bottom of the timeline, the user is able to explore the data freely, while still maintaining a clear context in regards to the surrounding data points. The clear disadvantage posed by using a 2D method, is that it does not afford the same ability as a 3D model to represent the same amount of information in the same space. This disadvantage is mitigated to an extent through the use of zooming functionality, and the ability of the end user to filter the events by type, or indeed by searching for text of their choice in the title or description fields. This approach provides the user with the full feature set as mentioned by Schneiderman in his “Visual Information Seeking Mantra” (Schneiderman, 1996). The “overview first” stage is presented to the user as

a timeline filled with all event data; at this stage each individual point of information may not be entirely clear to the end user. The next stage of the mantra, “zoom and filter” is catered for by the respective functionality built into the TimeGlider.js library. The end user is free to move around the timeline, zoom to a specific period of time, and to filter the events as previously discussed. Finally, the “details-on-demand” stage is catered for by the ability of the end user to click on any of the events and see more detail about the event.

However, although the format of the visualisation itself was promising in respect to the ability for different type of event information to be conveyed to the end user, the tool itself suffered from a number of critical problems which would preclude its use as a viable tool in a digital forensics investigation. These problems were largely related to the platform of the tool being a web based. Although providing a number of benefits, which were previously discussed, such as the ability to easily update the software; it also provided a number of significant limitations.

Whilst developing the prototype tool, problems were encountered with the security features of modern web browsers. Specifically, as the prototype tool was initially developed to run solely on the local device without the intervention of a web server, when required to load information from a JSON file the web browser would block this attempt. As a security precaution, the web browser will block any attempt to load these files unless they are requested from a web server using HTTP. This was simple to fix, as a small web server was loaded onto the local machine and used to serve the application. This would not pose a problem itself, as in a team environment it would be likely that the application would be hosted on a central server. The problem arises when the JSON file is extremely large. Not

only would this take some time to download from the server, but as the web browser tries to process the file, it can begin to encounter problems. During testing, when the John Doe test dataset was loaded, the web browser appeared to freeze for some time, and eventually repeatedly informed that the script had become unresponsive and prompted for it to be terminated. If the user allowed the script to run, the timeline would eventually load, however, when scrolled or zoomed to a different period in time, the application would again lock up whilst it attempted to internally process the event data. This combination of problems led the application to be entirely unusable. Adding further event data to the timeline, or allowing for significant interaction with the Autopsy database would likely exacerbate these problems further.

4.4.3 Case Study Outcomes

The main intention of this case study was to examine the advantages and disadvantages of moving away from 3D visualisations, and moving to a 2D format. Specifically, the prototype would indicate whether the common problems of obscured information and difficulties in pattern recognition which were encountered in the 3D case studies would also be present in 2D visualisations. From this case study, we can see that the problem of obscured information is still present in part, as a large number of data points need to be represented in a small space. However, this problem is mitigated by providing the user with tools to allow zooming and filtering of the dataset. Arguably, these features could have been developed in the previous 3D case studies; however, this would not solve the problem of distorted patterns depending on viewpoint. In the case of the Walrus case study, it would be problematic to implement a useful filtering tool without creating further problems. As the visualisation itself relies heavily on the links between nodes; by filtering out nodes, this would make these links difficult or impossible to see. Highlighting nodes using a different colour could be a potential

approach, however, this could pose problems if the ability to colourise the nodes based on an attribute has been utilised.

Overall, the format of the visualisation appeared promising based on this case study, due to the ability to display a large number of different event data points to the user. The timeline model is a format which is well-known and familiar to end users, suggesting that it may allow the user to examine the information and use the tool with very little prior training. This is an ideal scenario, as it would allow investigators to quickly start using the tool, and to start drawing conclusions from the dataset without the need to refer back to user guides.

The JSON input format of the tool is ideal, as it provides a standardised way to provide the tool with a summarised index of the dataset. Should a new event type be required, only a minor amount of modification needs to be carried out on the pre-processing script to allow these new events to be added to the timeline. The prototype application, however, does not allow the end user to add their own events to the timeline. This could potentially be of use for investigators to mark external events on the timeline which they know to have occurred. For example, if the suspect was known not to be in possession of the device in question for a certain period of time, this period of time could be flagged on the timeline. Such a feature would allow the investigator to quickly analyse the effect these custom points have on patterns within the dataset, and allows them to share their information with collaborators in a visual format.

The limiting factor for this case study was the platform on which it was developed. In this instance, the limitations presented by developing a web-based application made it unrealistic

to continue development. When loading a realistic volume of data into the visualisation, it becomes almost entirely unusable due to user interface delays, and the repeated prompts by the web browser to terminate the script as a result of these delays. It is suspected that there could be ways to reduce these delays is possible, however, as a prototype tool it would be unrealistic to pursue this further. Not only would the delays be frustrating to the end user, but it would undoubtedly cause bias in when testing this software to validate the hypothesis that visualisation methods when applied to digital forensics datasets can increase the effectiveness of investigations. Delays caused by the software would be perceived negatively by the end user and would increase the length of the investigation, thus actually decreasing the overall efficiency of the investigation.

From this case study, the conclusion was drawn that the timeline paradigm could be beneficial to the end user in a digital forensics investigation, as it provided the ability to display a large amount of chronological information to the end user through the use of its zooming and filtering features, whilst still allowing the user to maintain a clear perception of context. It also allows for patterns to be recognised in the event points, and conclusions drawn from these. However, the aforementioned performance issues encountered would preclude the ability to conduct meaningful experiments with a group of end users. As such, this case study provides a well-structured framework, with elements such as the index file aspect drawn from the previous case studies; which allows for the development of a prototype tool which would be suited to final experimental studies.

4.5 Conclusion

Through these case studies a number of lessons have been learned. Initially it was envisioned that 3D visualisation method would present an ideal way in which to present the large complex datasets which are used in a digital forensics investigation. However, the first case study showed that when presenting log data which is primarily chronological in nature in a 3D space, it is difficult to consistently identify patterns in the dataset. This is due to the fact that the data points are not hierarchical in nature, and so are not explicitly linked. When the user rotates the model to view it from a different perspective, any links which they had mentally drawn are distorted. The application was also very slow to load as it had been designed to directly interact with the Autopsy SQLite database, querying for information every time the application was loaded. As the database never changes, this is unnecessary. As a solution to this problem, it was decided that an index of relevant data should be created for use in future case studies.

The second case study took the difficulty of identifying patterns in 3D space into account and instead looked at data which is hierarchical in nature, and thus explicitly defines links regardless of the user's perspective of the model. This was successful in showing file system structures, with information such as timestamps represented by colours. While this was successful, it was deemed that this information alone was not of enough use in a digital forensics investigation, and that information represented in the first case study (log files, browser history, important files etc.) would be of more use to the end user. This case study also utilised the data index method identified in the first case study. This was explicitly required by the application as it used its own file format to draw graphs.

The third case study reverted to looking at chronological event data, as the previous case study showed that filesystem data alone was not enough to allow the end user to draw conclusions about user behaviour. Having identified a difficulty in displaying chronological information in a clear way using 3D techniques, it was decided that the third case study would use 2D methods instead. Drawing from the work of Lowman and Ferguson (2010), the idea of a timeline which would show a number of data points from the dataset was used. This would allow the end user to view event data, as was used in the first case study. An index was built of relevant data in JSON format, which was then loaded into the prototype application which had been developed as a web based application using JavaScript and D3.js. The format of this application was promising, as it allowed the user to explore the dataset in a familiar format, and allow them see patterns of behaviour of the suspect user. However, problems were quickly encountered regarding the performance of this tool. Not only did it take some time to create an index, but when this index was loaded into the application, it would often freeze, and browser safeguards would kick in, announcing that the script had become unresponsive. These browser problems led to the application becoming completely unusable when using a dataset of any realistic size.

Chapter 5 – Insight Prototype Development

5.1 Introduction

The outcomes from the various case studies identified that a promising format for the prototype would be a 2D timeline, as this would allow various types of chronological information from the source dataset to be mapped to this timeline and displaced to the end user. This would allow the user to view the activity of the user in regards to their web browsing, application installations, creation of EXIF tagged photos etc. The case studies showed that 3D methods were not well suited for this purpose, and that the application would be best developed to target the desktop environment due to the restrictions imposed by web browsers. A pre-processed index format was also to be implemented to address the problem of slow loaded identified by the first case study.

5.2 Insight

5.2.1 Prototype Development

The purpose of the final case study was to draw together the outcomes of the previous case studies, and use the lessons learned to build a prototype tool. This prototype tool would be used to validate the hypothesis that utilisation of exploratory information techniques can increase the effectiveness of digital forensics investigations. From the previous case studies, it was concluded that the format of the final tool would be a 2D timeline tool which would contain various types of event data from the Autopsy database. Event data would be summarised in a human and machine readable format, such as JSON, in order to expedite the loading of the application, and to allow for new types of event data to be added to the application with ease. The application would also interact with the Autopsy database to retrieve additional information as it was requested by the user.

This case study looked largely to build on the outcomes of the previous ‘Insight.js’ case study, by adding functionality and addressing issues which were encountered. The core of these issue is the performance aspect of the web application. It was found that a web application was not an ideal platform for processing and interpreting datasets of a realistic digital forensics scale. As such, the prototype tool for this case study would be a desktop application built using a well-developed framework. In this instance, an application written against the popular Microsoft .NET framework was decided to fit this requirement. The .NET framework (Figure 5.1) provides a mature and robust, enterprise-grade framework, against which Windows desktop applications can be built. For this reason, it could be assumed that applications developed against this framework would handle large datasets with ease, and would be executed without restrictions as was found during the development of the ‘Insight.js’ web application.

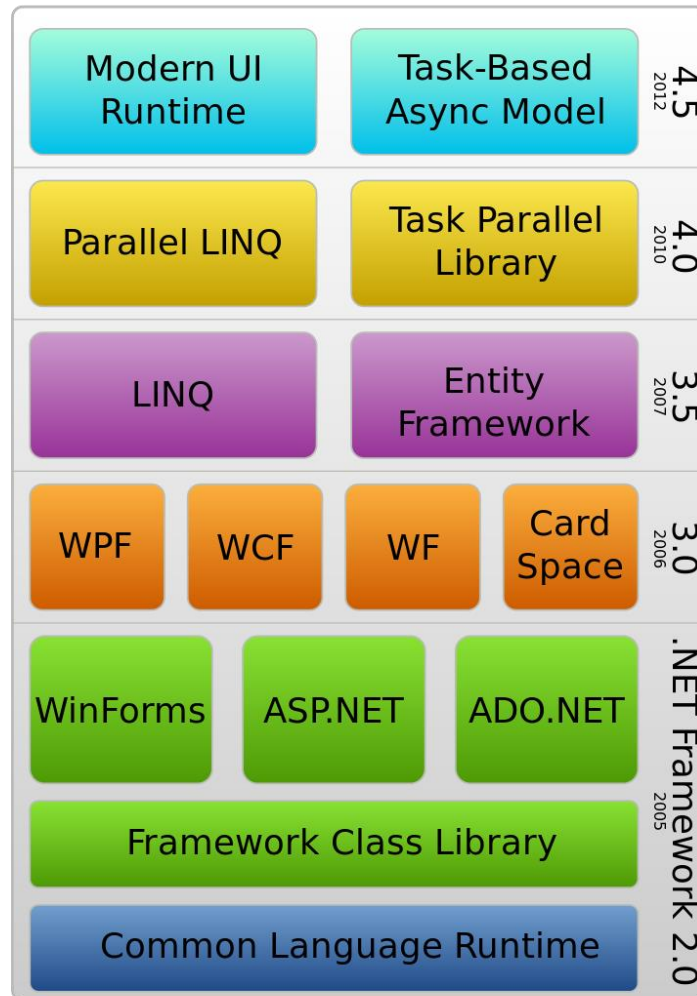


Figure 5.1 - .NET framework structure ("DotNet" by Soumyasch, Wikimedia)

Such a robust framework also provides the developer with a significant number of libraries to develop against. This large set of standard libraries allows most of the functionality of the application to be written without the requirement to find libraries on the internet written by third-parties, which may not be as well-tested and as such could potentially introduce performance bottlenecks into the application. Libraries which are not included as standard are often available as an optional download within the development environment from either Microsoft or trusted third-parties. An example of such a library is the SQLite library which allows .NET applications to interact with a SQLite format database. This library was provided by the developers of SQLite, and is one of the libraries which was downloaded in

order to allow live access to the Autopsy database. The .NET framework also allows developers to choose from a number of different languages, such as C#, Visual Basic, F# etc. When compiled, the developer's code is converted to Common Intermediate Language (Microsoft, 2016). This feature means that libraries which have been written by a developer in one language, they can be used by another developer writing an application in any language which compiles to the Common Interface Language. This is a major benefit, as it allows the expertise of different developers to be pooled, regardless of the language they are most experienced with. Such a benefit would allow any prototype tool to be further developed by the community if ever publicly released, or would allow a comprehensive plugin system to be developed. In compiling to Common Intermediate Language, it also means that the application is technically platform-agnostic. That is, the application could run on an operating system other than Windows, if a compatible framework is present. Such a framework exists for Linux, called Mono (Mono Project, 2016). However, the Mono framework is not 100% compatible with .NET applications (<http://www.mono-project.com/docs/about-mono/compatibility/>) and has some functionality missing or incomplete. As part of this research, the prototype application would not be tested or developed with the Linux operating system, although it could feasibly be adapted to work.

There a number of different ways in which the .NET framework can be utilised in regards to how the applications are rendered by the system. The two most prominent options available to the developer are the older, but still widely utilised WinForms system; or the newer, more flexible Windows Presentation Foundation (WPF). Due to its flexibility in regards to the display of rich graphics and UI elements, WPF was the chosen system for the development of the Insight prototype application. The WPF allows for a user interface to be developed using the Extensible Application Markup Language (XAML); an XML-based language which is

used to declare the structure of the user interface, and attributes applying to each element; for example, the size of an element, its colour etc. Such an approach separates the application logic from the user interface design, allowing changes to be made the user interface design, whilst avoiding the need for any substantial changes to the main application code. It also allows the developer to clearly group user interface elements, and to optionally bind them to a dataset.

As a comprehensive and popular framework, this also meant that one of the libraries available to install from within the development environment is a timeline control (Syrov, 2012). This control allows the developer to use a fully functional 2D timeline in their application, without the requirement to build the timeline from the ground up. The timeline framework provides elements such as scrollable sub-timelines which allow the developer to design a timeline which may, for example, have a large main timeline, showing 24 hours of event data, below this main timeline, there may be a smaller timeline which shows 2 weeks of data (in the format of data point, omitting detail). Below this, there may be another smaller timeline which shows an even wider range of time, for example 2 months. This approach gives the user a strong overview of the data and allows them to retain a solid context. It also substantially eases the process of jumping to a different point in time. For example, if the user would like to view a day that with 2 weeks in the past, instead of having to scroll the main 24 hour timeline for a while, they could use the smaller 2 week timeline to scroll. When the user interacts with one of the timelines, all of the others scroll respectively. The framework also provides a simple way to load data into the timeline through the use of a XML file. Such an ability to load from an XML file allows the index structure which was developed in the previous case study to be used. However, as the format used in the last case study was JSON, the pre-processing script required modification to instead output to XML format. This

modification related largely to the formatting of the file, as the method in which the data was retrieved from the Autopsy database and the data itself was mostly unchanged. In order to consolidate the application to a single language, the pre-processing script was ported use the C# language and the .NET framework. As the .NET framework provides a native ability to interact with and generate XML files, this was a relatively uncomplicated exercise. In porting the pre-processor script to the C# language, it also would allow the functionality of this to easily be integrated with the main Insight application to provide an all-in-one solution. However, for the purpose of this research, the pre-processing functionality was maintained as a separate application for clarity.

The initial stages of the development focused on developing a basic user interface which provided a timeline visualisation to the end user, which would load all event data from the relevant XML file at load time and display it to the end user. It was decided that in order to assist the user's ability to efficiently explore the dataset, the timeline visualisation should have one main timeline which displayed all of the events in the dataset, along with a summary of the event. This timeline would be supplemented with 2 sub-timelines which would provide the user with a 5 day and 5 month overview of the dataset respectively by default (Figure 5.2). By giving the user a view of different chronological periods at once, this can assist the user in exploring the dataset rapidly. This can be very important if the user is investigating a dataset which spans a very large period of time, such as a number of years. The application also provides zoom functionality, which allows the end user to alter the period of time which is shown. Therefore, if they are not satisfied with only being able to see 5 months on the bottom timeline, they need only zoom out, and all timelines will adjust their scales automatically.

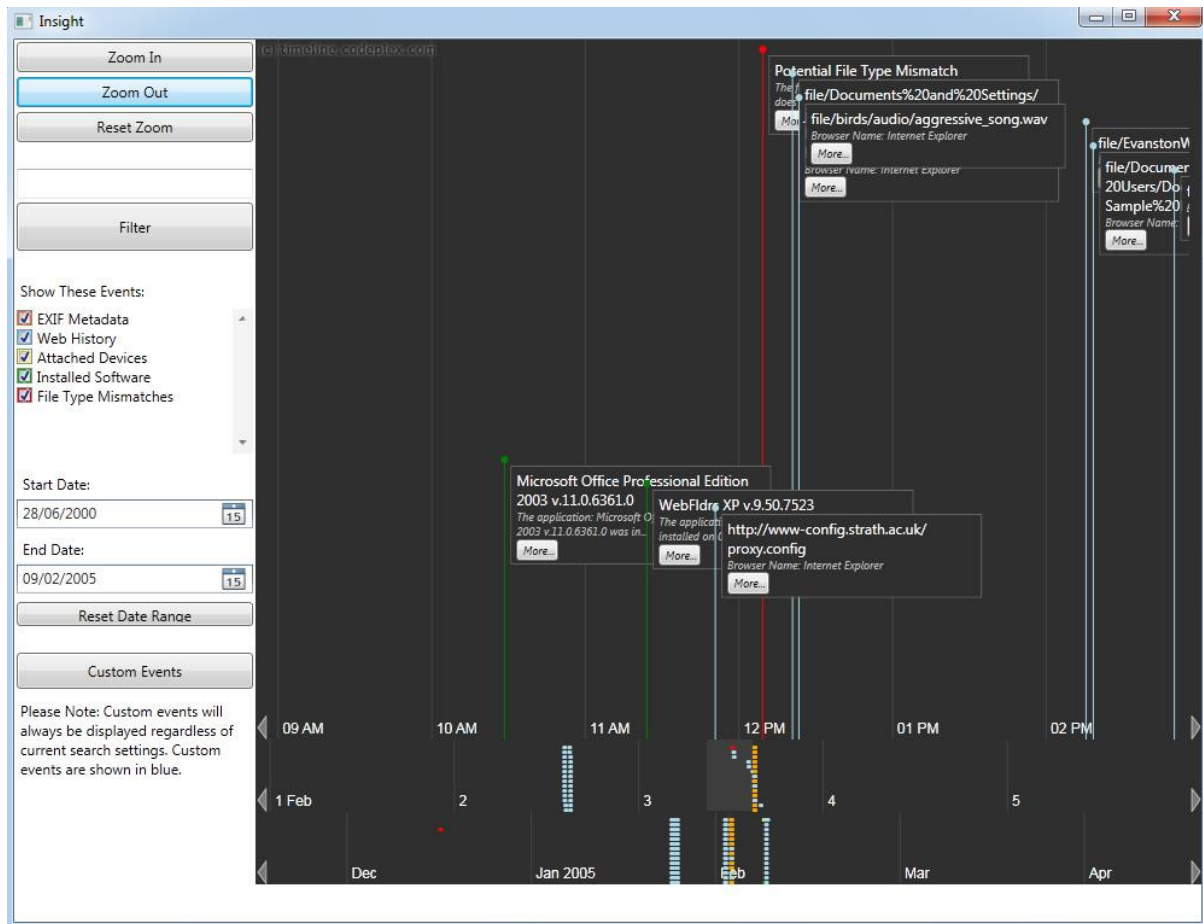


Figure 5.2 – Insight showing 3 timelines with different time scales

Each event on the timeline displays a title which shows basic information about the event at a glance, and a small description. The application was developed so that when the user clicks on the description of the event, a small popup appears giving them more information about the event. As one of the supported types is image EXIF metadata, when the user clicks on one of these events, the detail popup will attempt to load the picture from the dataset (Figure 5.3) and display it to the end user. In doing so, the end user is not only afforded the ability to explore the events surrounding this image, but also to view the image itself without the need to leave the Insight application.

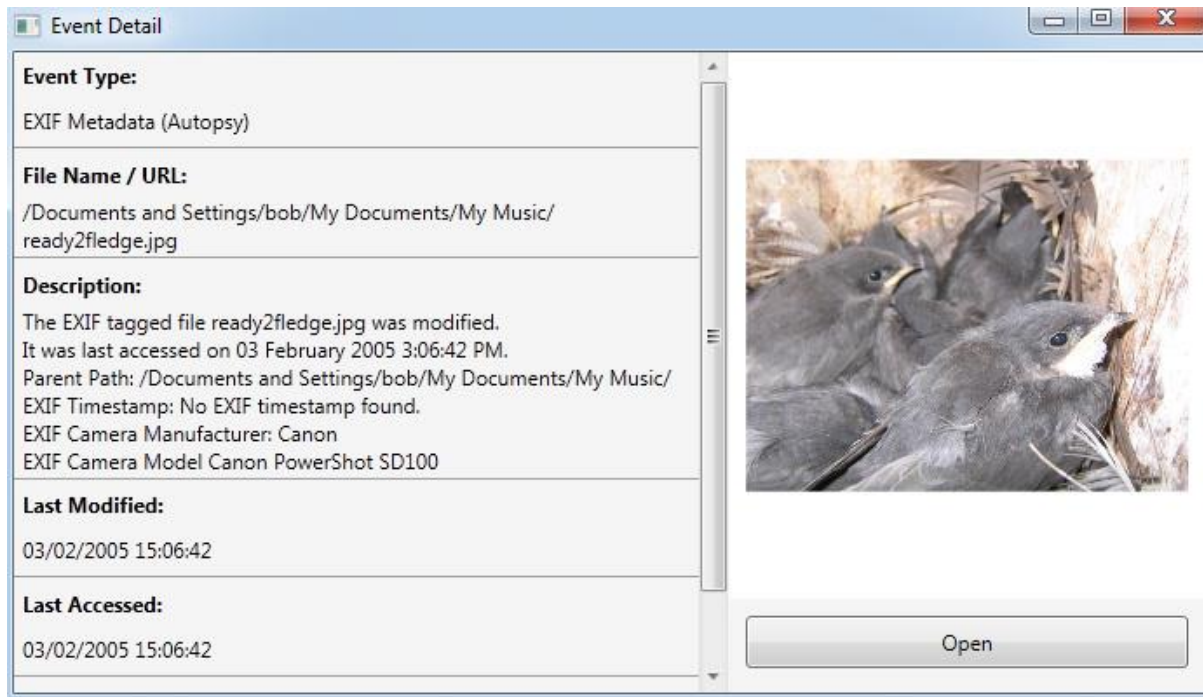


Figure 5.3 – Insight event detail view showing image file from source disk

The timeline itself was developed to the stage where all relevant data could be displayed on the timeline with the ability to draw additional information from the Autopsy database, and where relevant, from the source file system. Once this was complete, additional functionality was developed to allow the user to filter data by either event type, or by searching for text (Figure 5.4). As this functionality was not integrated into the timeline library itself, it involved reading the internal collection of events held by the timeline, creating a temporary copy of these, and then filtering the collection. By maintaining a copy of the original event collection, it ensured that when the user disabled any filter, that the original events would quickly be restored to the timeline without the need for them to be reloaded again from the index XML file. Although this approach increases the memory utilisation of the application, it ensures that the performance of the application is not impacted by having to read the entire index from disk every time the user changes the filtering criteria.

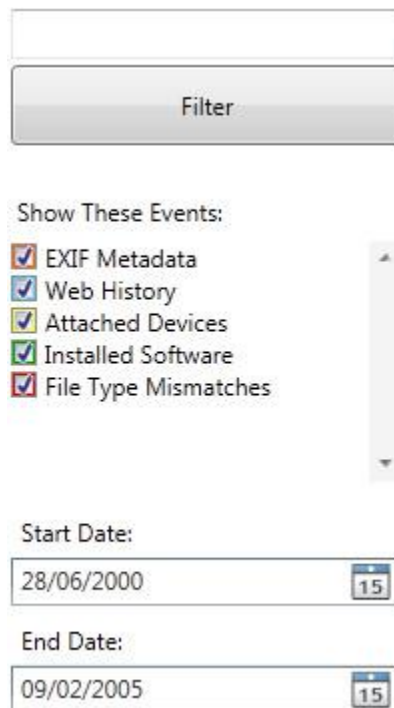
The image shows a user interface for a filter panel. At the top is a text input field. Below it is a button labeled "Filter". Underneath the button is a section titled "Show These Events:". This section contains a list of five items, each with a checked checkbox and a colored square icon: "EXIF Metadata" (red), "Web History" (blue), "Attached Devices" (yellow), "Installed Software" (green), and "File Type Mismatches" (red). To the right of this list is a vertical scrollbar. Below the list are two date selection fields. The first is labeled "Start Date:" and contains the text "28/06/2000" with a calendar icon showing the number "15". The second is labeled "End Date:" and contains the text "09/02/2005" with a similar calendar icon showing the number "15".

Figure 5.4 – Insight filter panel

Additionally, the previous case study identified that it could assist the end user in exploring the information if they were provided with the functionality to mark their own events on the timeline. They may choose to do this to mark narratives of behaviour they have discovered, to share notes with colleagues, or even to integrate external events into the timeline. In order to develop this functionality, it was necessary to develop a method in which the user would be able to provide relevant event data to the application, which would then be formatted in a way that could be understood by the timeline element. This was a straightforward process, as the timeline library stores each event as a new `TimelineEvent` object. This type object has a number of fields which store information about the event such as its title, timestamps etc. When the user adds an event to the timeline, the details they have entered can be used to create a new `TimelineEvent` object, which in turn can be added to the timeline using its

provided methods. However, this approach meant that when the user closed the application, all of their custom events were lost. In order to solve this problem, it was necessary to create a class called `SerializableEventHelper`. This class would take a timeline event, extract the core information, and package it into a serializable class, in this case called `SerializableTimelineEvent` (Figure 5.5). All of the events created by the user would be converted to this new format, and then would be written out to an XML format file called “customEvents.xml”. When the Insight application is loaded, a method from the `SerializableEventHelper` class is called to reverse the process. All of the `SerializableTimelineEvent` objects are deserialized from the XML file, the information extracted from each of these and used to create `TimelineEvent` object. These are then given to the timeline element to display.

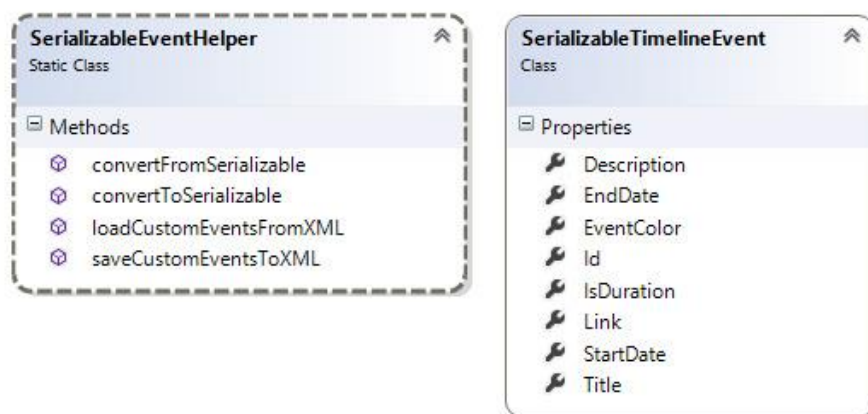


Figure 5.5 - Class Diagram

By continuing to use XML as the format for custom events, this also provides the end user with the option to modify the file outside of Insight should they wish to do so. It also affords the ability for other tools or scripts to write this file format should they wish to add custom events to the Insight timeline. These custom events could have been integrated into the main event XML file; however, in maintaining a separate custom event file it allows the end user

to use these custom events across multiple datasets. A scenario in which this could be useful is when the investigator may have datasets from numerous devices but which are all involved in the same case. The investigator can load one dataset and create custom events against this to note their findings. They then have the option to load a new index XML file for one of the other datasets in the case. However, they may still preserve the existing custom events XML file. When the Insight application loads, the user will be presented with a timeline visualisation of the new dataset, but will still see the custom events they have created previously. This approach can allow an investigator to preserve a sense of context across multiple datasets, and affords them the ability to build up a visual model of their understanding of the case as they progress through the datasets.

Interestingly, the timeline framework also provides the feature of inertia scrolling for the timeline. This means that the user can “flick” the timeline to a side, and it would use the “momentum” of the flick to continue scrolling for a short period of time. This feature is now most commonly seen utilised on smartphone operating systems. It was decided that this feature would be turned on, as it would provide a familiar user experience to people who are used to having this functionality.

5.2.2 Preliminary User Testing

Initial testing of the application was in the format of small pilot studies undertaken by the researcher. The purpose of these pilot studies was to assess the basic functionality of the application, and its ability to handle large datasets. As the performance of the prototype application from the previous case study was a significant issue, this was under constant scrutiny during the development of the Insight application during this case study. In regards

to the pre-processing application, this took a much shorter time than in previous case studies; in many cases only taking a few minutes to process the dataset.

In regards to the main Insight prototype application, when loading the index XML file, this also was much faster than in previous case studies. The John Doe test case presented no noticeable delay in loading the application, and the realistic scale dataset only produce a very minor delay of a few seconds. However, as was expected, the memory utilisation was substantially higher when loading the larger dataset. Due to the low cost of RAM and the large amount often found in modern PCs, this was still deemed to be well within an acceptable range, as it did not exceed 1GB of RAM utilisation at any time.

These early case studies by the researcher indicated that this prototype successfully integrated the outcomes of the previous case studies, and had the potential for use as a tool to validate the hypothesis of this research. In order to ensure that the prototype application was at a suitable stage to conduct full experimental analysis, a pilot study was first conducted. This pilot study was conducted on a small group (n=5) of third year Ethical Hacking students. Although these students had no extensive training in digital forensics, this was not a problem, as the pilot study focused solely on the usability of the application. Each participant was given a set of exercises to conduct using the Insight application. These exercises were designed so that the participant would use most of the functionality of the application. Qualitative feedback was gathered from the participants at the end of each exercise, and at the end of the session regarding their experience with using the tool in general. The purpose of this case study was not to gather statistical evidence to validate the hypothesis of the research at this stage; rather, it was used as a tool to test the application with users, and to assess the

usability of the application. From this pilot study, a number of errors were encountered by the users during use which had been overlooked during development. The feedback from the users was used to correct these errors, and to improve usability in areas of the application which had been identified by the participants. Although the participants were asked to provide answers to questions regarding the dataset during the study, the correctness of their answers was not examined. At this stage this was unnecessary, and results based on the accuracy of their answers would provide little value to this research, as the participants did not have the required level of digital forensics training to act as a reliable user group for this purpose. This pilot study also provided an opportunity to examine the use of a virtual machine environment during user tests. In order to provide each user with an identical environment with the required software and pre-processed datasets in place. In using a virtual machine environment, it ensures that as each environment is identical, any results derived from experiments conducted within these environments are sound and reproducible. This pilot study successfully implemented a virtual machine, as none of the participants experienced any problems when using it. As the study was conducted on a single laptop, it also allowed the environment to quickly be rolled back to a preset state between each participant. It was, however, noted that this was a manual process and any future user studies conducted using a virtual machine would benefit from this step being automated.

5.2.3 Validatory Testing

Once the pilot study had been completed, and the Insight application modified to incorporate the feedback from users; the next step was to conduct a full user group experiment in order to validate the research hypothesis that through the use of exploratory information visualisation techniques, the effectiveness of digital forensics investigations can be increased.

In order to test this hypothesis, a user study to be conducted on a group of students (n=29) studying Digital Forensics at Masters and 4th year undergraduate level was proposed. Such a group would allow the prototype application to be tested by a number of people who all possessed the required knowledge to understand the process of a digital forensic investigation, and who all had a similar level of experience and skill. This user study presented the participants with 6 questions to answer regarding the John Doe test case. These questions would be answered using one of two tools. The first of these tools was the prototype Insight applications, and the other Autopsy 3. The use of Autopsy 3, as a primarily text-based tool, would provide a control against which potential effectiveness gains, and user experience perception could be measured. This was assessed to be a suitable control tool as the participants had not been instructed during their course on the use of Autopsy 3, thus would likely be using either of the tools for the first time. The tasks and questions (Appendix A) were designed with the intention that the participants would be required to make use of most of the features available in the application, such as the filtering and event detail features, and were designed to reflect the type of information an investigator would be looking for in a real investigation. This includes information about the photos on the suspect device, where the suspect has gone in the physical world based on information from their device etc. Such information would be useful in operations such as Innocent Images (FBI, 2006).

Each participant was given a task sheet at random, each marked with a random participant identifier. 14 of these sheets indicated to the participants that they were to use only the Autopsy 3 tool to complete the experiment, and the remaining 15 indicated that the participant should complete the tasks using the Insight tool. In order to attempt to reduce the potential for participants sharing information about the experiment with others working beside them, they were asked not to discuss their results with each other during the

experiment. Each participant was instructed to use a virtual machine environment which had been distributed to all lab computers in preparation. This virtual machine was pre-loaded with versions of Autopsy 3 and Insight installed, along with pre-processed datasets to allow the participants to start examining the case immediately. As had been recognised in the Insight pilot study, resetting these virtual machines manually would be a laborious task with such a large number of participants. In order to simplify this process, all of the hard drive files of the virtual machines were set into “non-persistent” mode. This means that when the virtual machine is powered on, the participant can use it like they would a normal computer. However, when it is powered off, all changes made to the hard drive are discarded and it reverts to its previous state. This approach ensured that when powering on their virtual machines to start the experiment, all participants had an identical environment to work within.

As part of testing whether there are any gains in effectiveness, one of the metrics used was the efficiency of the investigation; that is, how quickly an investigator completes the investigation. In order to measure this, it was necessary to time the participants. By timing how long they took to complete the experiment, this can give an indication as to whether there are likely to be any efficiency gains when applied to a full scale digital forensic investigation. This posed some problems, as it would be impossible to manually time all of the participants. The participants were required to all complete the experiment in the same session to ensure that they did not share information about the answers with people outside of the session. Had multiple sessions been conducted, and information shared between participants, this could have led to participants completing the experiment in an artificially short time, thus skewing the results. Also due to the inherent shortage of available digital

forensics datasets, it would not have been feasible to hold multiple sessions and give each session a different dataset.

In order to solve the problem of timing the participants, a small application was developed which would allow the participants to be automatically timed. This was written as a C# application which the user was instructed to run when they first accessed the virtual machine (Figure 5.6).

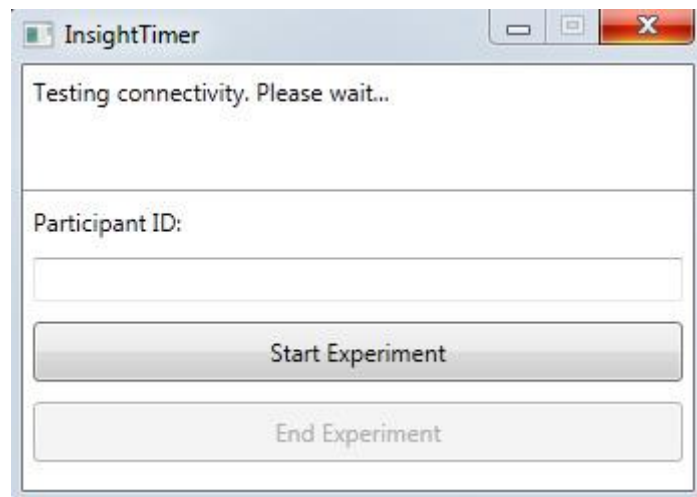


Figure 5.6 – Experiment timer application

The application would conduct some network connectivity tests to ensure that the virtual machine had internet access. If these tests passed, the participant would be prompted for the participant ID which was given to them at the top of their instruction sheet. Once they had read the instruction sheet fully and were ready to begin the experiment, they were asked to click the “Start Experiment” button on the timer application. When this was clicked, the application would send a simple POST request to a web API which had been developed and

was running on an external server. This POST request contained the unique participant ID, the event type (either a start or stop) and a preset API key which was used to validate the request. The web API then took this participant ID, and inserted it into the MySQL server which was running on the server, along with the server timestamp, and the event type. The participant was asked to click the Stop Experiment button once they had completed all of the tasks, which would again record this time on the web server, allowing the time taken by each participant to be derived. At the end of the experiment, the web API was modified to accept a different API key. This was done to ensure that if anyone used the timer application outside of the experiment, the web API would recognise the change in API key, and would send a reply to the timer application stating that the request had been rejected. The timestamps stored by the application are taken from the database engine itself rather than received from the timer application. As such, the accuracy of the clocks on the virtual machines is irrelevant, as all times are taken from a central source. Hosting the web API on an external server also protects the results of the experiment, as it is off-site and backed up regularly.

At the end of each task, the participants were also asked to rate how easy they found the task to complete with the given tool. They were given a 5 point Likert scale with a value of 1 representing “Very Difficult” and a value of 5 representing “Very Easy”. In collecting these ratings, it could be assessed if participants displayed a strong preference for either of the tools on a per-task basis. As each task targeted different types of event data, or different approaches to analysing the information, it would give a potential indication into whether either tool was found by participants to be particularly suitable for solving certain types of problem.

It should also be noted that the participants were all asked to indicate on their instruction sheet if they were colour-blind. Due to its use of colour to represent different event types, colour blind users would likely find the application especially difficult to use. The privacy of the users was protected regarding this matter, as no personal information was collected on users. Any colour blind participants were identified solely by their random participant ID. They were still allowed to participate in the experiment, and were asked to indicate in their answers if there were any parts of the experiment they struggled with because of their colour blindness. When analysing the results of this experiment it was found that one participant had identified themselves as suffering from colour blindness. However, as the participant task sheets were distributed randomly, this participant had been assigned to the control group using the Autopsy 3 tool. As this is a primarily text based tool, the participant's colour blindness is likely to have had minimal impact on their results.

5.3 Conclusion

From this case study, qualitative results were collected from the feedback provided by the participants in the user study, regarding their experiences using the allocated software tools. Quantitative results were also collected in regards to the time taken by each participant, their Likert scale responses, and the accuracy of their answers. These results will be analysed in detail in the next chapter.

Chapter 6 -Results

6.1 Introduction

As was discussed in the last chapter, a group (n=29) of students studying Digital Forensics at Masters and 4th year undergraduate level took part in a user study with the aim to validate the research hypothesis. The hypothesis is that by utilising exploratory information visualisation techniques in digital forensics investigations, gains in effectiveness can be realised. Effectiveness for these purposes is defined as a combination of the efficiency and accuracy of the investigation. We can say that in general, an investigation is more effective if one or both of these metrics improves.

The user study involved presenting the participants with a variety of tasks to complete which involved finding information and answering questions regarding the John Doe test case. Each participant was allocated either the Autopsy 3 or Insight tool before starting the study, and were instructed to only use this tool to find the solution to the tasks. As a primarily text-based tool, Autopsy 3 represents the typical format of tool used during the analysis phase of the digital forensic investigation lifecycle. For this reason, it was a suitable control measure for the user study. The metrics derived from the user study in terms of time taken, Likert ratings, and overall accuracy of answers were compared between the Autopsy tool and the Insight tool. This approach allowed the research to determine if there were any clear advantages of using the Insight visualisation tool, over the control Autopsy 3 tool. These results were used to validate either the research hypothesis as discussed previously, or to validate the null hypothesis. The null hypothesis for this research is that the utilisation of exploratory information visualisation techniques in digital forensics investigations provides no benefit in regards to effectiveness.

6.2 Presentation of Results

6.2.1 Completion Time

The primary metric used by this research to measure the efficiency of the investigation is the time taken to complete all tasks. As was discussed in the previous chapter, the participants were timed using a tool which recorded their start and finish times. The participants were instructed to start the timer just before they began the tasks, and then to stop the timer once they had completed the task, but before they completed the general feedback section of the experiment.

Of the 29 participants, 14 were allocated to use the Autopsy tool and were given a participant ID prefixed with “AA”. The other 15 participants were allocated to the Insight tool and given a participant ID prefixed with “IA”. Of the Insight participants, all 15 completed the experiment. However, of the Autopsy participants, 11 completed the experiment. The reasons for this were: 1 participant (AA48115) recorded a start time but failed to record a finish time, 1 participant (AA55372) recorded neither times, and 1 participant (AA30846) had to leave before they had finished the experiment. Any feedback and Likert scale results from these participants will be included in the relevant results, but the participants will be excluded from completion time and accuracy rate calculations as they provide incomplete data. Before the experiment began, all participants were informed that although their times were being recorded, there was no time limit on the experiment. It was important to convey this point so that the participants did not feel rushed. Imposing a time limit would be likely to have the effect of artificially decreasing the completion times, whilst simultaneously decreasing

accuracy rates. It also allowed the participants to become accustomed to the tool at their own pace, thus giving a more realistic reflection of a real-world investigative scenario.

The individual timestamps were retrieved from the MySQL server at the end of the user study, and were used to calculate the duration of time taken for each participant to complete all tasks (see Appendix E). Based on these results, a mean time to completion for the Autopsy participants (n=11) can be calculated as 1 hr 5 mins (3922 sec). The minimal and maximal durations for Autopsy participants were 00:37:14 and 02:22:08 respectively. The mean time to completion for Insight participants (n=15) is calculated as 56 mins (3372 sec). The minimal and maximal durations for Insight participants were 00:23:15 and 02:02:38 respectively (Table 6.1). In respect to the mean completion time from each group of participants, it can be seen that the Insight application provides an average of a 14.02% reduction on completion time when compared to the Autopsy 3 tool.

However, the p-value for the two sets of duration results when calculated using a two-tailed unpaired t-test is 0.4382. As such, although a reduction of the average completion time can be shown, when a significance value of 0.05 is used, the calculated p-value shows there is no statistically significant difference between the completions times of the Autopsy and Insight test samples. This metric alone therefore does not allow us to reject the null hypothesis.

	Autopsy	Insight
Min	00:37:14	00:23:15
Mean	01:05:22	00:56:12
Max	02:22:08	02:02:38
p-value	0.4382	

Table 6.1 – Completion Times

6.2.2 Accuracy Rates

It is not sufficient, however, to simply analyse the time taken by each participant to complete the investigation. If the participant was able to complete the investigation in a faster time, yet produces answers which are largely incorrect, then it can be said that the tool itself is flawed and is not fit for purpose. The accuracy of the investigation should under no circumstances be compromised for the sake of a faster investigation. This is especially important when the investigation is for the purposes of law enforcement, in which inaccurate information can have significant consequences. For this reason, the accuracy of the participant's answers was assessed to ensure that the participants were able to complete task correctly, and that this compares to a similar accuracy level as the control tool, Autopsy 3. Each question was assessed individually as the type of information involved in each differed. It would be unwise to combine the accuracy rating for each question into an 'overall accuracy' measure, as it cannot be assumed that each question is equally weighted. Also, in instances where a task is composed of a number of sub-components, the task as a whole is only considered to be correct if all of its sub-components are correct. Where the participant provided no response for one or more components, the task is also considered to be wholly incorrect. The accuracy ratings, i.e. the percentage of wholly correct responses can be seen in Table 6.2 and Figure 6.1. A breakdown of these ratings is provided as Appendix B.

	Q1	Q2	Q3	Q4	Q5	Q6
INSIGHT	80.00%	100.00%	46.67%	100.00%	80.00%	80.00%
AUTOPSY	54.55%	90.91%	72.73%	100.00%	9.09%	72.73%
p-Value ($\alpha = 0.05$)	0.1730799	0.4230769	0.1925642	1	0.0004598	0.6698151

Table 6.2 - Accuracy of responses to each task in user study

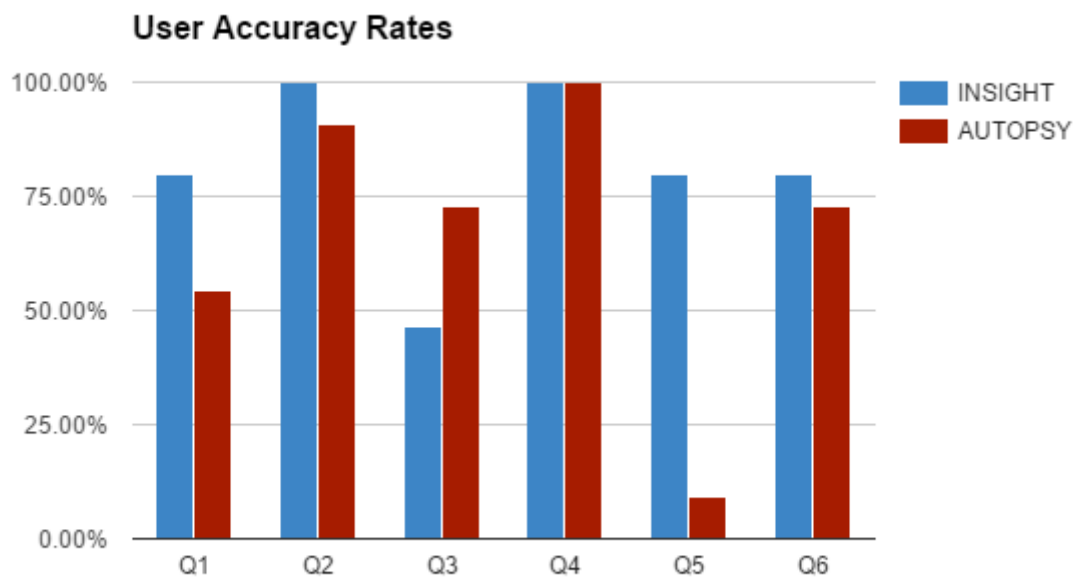


Figure 6.1 - Comparison of accuracy rates between participants using Insight and Autopsy

From these results, it can be seen that the accuracy ratings for the Insight tool are slightly more favourable in most tasks. In most cases, the accuracy rating is equal to or better than the Autopsy 3 tool. Question 1 asked the participant to find when the suspect installed the Firefox web browser software on the device, and from where it was downloaded. 80% of Insight participants answered all components of the task correctly, whereas only 54.55% of Autopsy participants answered all components correctly. The most common error that was made was incorrectly stating the download source. In order to answer this question, the participants were required to examine the web history data, and would likely use the search functionality of the application. This question would give an indication of how the participants would cope when asked to find a known event (in this case the installation of Firefox), and to examine the events in close chronological proximity. By skipping any simple “training questions”, this can also provide us with an idea of how well the participants cope with each tool without any prior tuition. It was expected that the Likert scale results and the feedback provided by participants would provide more insight into how the participants felt when using the application for the first time.

Question 2 required the participant to find an image on the suspect system which had been taken with a Canon EOS-1D camera. They also required to give the time and date the file was last accessed. In order to complete this task, the participant would need to look at the EXIF metadata of various image files. Both tools showed similar high accuracy rates in this task. Insight participants scored slightly higher with an accuracy rating of 100%, and Autopsy participants also scored a high accuracy score of 90.91%. This task examined the participant’s ability to find a known event within a subset of the dataset. In this case, in asking the participant to find a photo taken with a certain model of camera, this would indicate to them that they would be required to look within EXIF metadata. This subset of

information is easily accessible from within both tools. Unlike Question 1, this task did not require that they examine chronological information. This task should be simple to solve if the participant makes use of the filtering functionality of the applications to show only EXIF tagged files. The task allows us to assess whether the participant is able to find additional information when given piece of external knowledge (in this instance, the model of camera owned by the suspect), regardless of any chronological information. This tests whether information can still be found using Insight even if the participant is not provided with a timeframe in which the event occurred, and how this compared with results from Autopsy participants.

Question 3 asked the participant to identify a zip file which the suspect had attempted to hide, and to name two of the files contained within this zip file. The attempt to disguise the file was made by changing the file extension of the zip file to .dll. This task was the only task in the user study in which Insight scored a lower accuracy score than the Autopsy 3 tool. In this task, the accuracy dropped to 46.67% for the Insight tool, whilst the Autopsy 3 tool maintained an accuracy rate of 72.73%. It was identified from participant feedback that many of the Insight users struggled to open the zip file. As this functionality was not integrated into Insight, they were required to use a third-party archive application which had been pre-installed onto the virtual machine. Therefore, they were unable to name the files inside the zip archive. Many participants, regardless of the tool they were using, mis-identified the 'hidden' zip file as another zip archive called 'birds.zip'. However, no significant attempt had been made to hide this file. This task was used to assess the ability of the participant to examine filesystem information when using Insight. In this case, the extension mismatch filter allows the user to find this information. However, they were then required to open the

file, which a number of Insight participants struggled with. As mentioned, this was reflected in the accuracy rate of Insight users.

Question 4 involves a single component which asks the participant to find the model of camera used to take a large number of the photos found on the suspect device. In order to find the answer to this question, the participant was required to examine the majority of the EXIF - tagged images on the device. No information was requested in regards to the timestamps of the photos which in turn tested the ability of the participants to find information which was not necessarily linked to a timeframe. For this task, both Insight and Autopsy participants demonstrated a 100% accuracy rate. This task was similar in format to Question 2, in that the participant was required to examine EXIF tagged files, however, in this case, the user was not given any external information. Instead they were required to search for the most frequently used camera model to take photos found on the suspect device, thus requiring them to examine data across the entire suspect device, regardless of time frame. This task allows us to test in a similar manner to Question 2, that the user can search the data without being guided by a timeframe. By presenting a similar task to Question 2, it would allow us to assess whether the accuracy ratings differed once the user had had more time to familiarise themselves with the application.

Question 5 informs the participant that the suspect used their device in a university at a point in time. The participant was asked to find more information about this, such as which university the suspect visited, how the participant knows this, and the date on which the participant visited the university. The format of this question was designed so that the participant is given a small fragment of external information, and is then left to explore the

dataset in an attempt to find corroborative information to support this external knowledge. In this task, the accuracy rates between the two groups of participants varied greatly. While Insight participants demonstrated an accuracy rate of 80%, the Autopsy participant group demonstrated an accuracy rate of only 9.09%. The exploratory nature of the Insight tool can allow the participant to see patterns of web browsing and as such, this may have simplified the process of identifying the beginning of a web browsing session, which started with the proxy configuration for the University of Strathclyde being loaded.

Question 6 specifically dealt with chronological information. The participant was instructed that the suspect had done something which could be classed as suspicious at 4:32pm on the 2nd Feb 2005. The participant was asked to give information about this event regarding what the suspect did at this time. This task was designed to give an indication of how well participants from both groups performed when given only a fragment of chronological information. This was intended to test the performance of the timeline paradigm in comparison to the common text based presentation of information as utilised by Autopsy. The accuracy rates in this task were slightly greater for Insight participants, with an accuracy rating of 80%. Autopsy participants were slightly less accurate, with a rating of 72.73%.

When looking at the accuracy rating results overall, it can be seen that in 5 of the 6 task presented to the participants, the Insight test group maintained an accuracy rating which was equal to or greater than the accuracy ratings of the Autopsy group. The results for Question 3 show that accuracy rating of Insight participants was 46.67% when compared with the 72.73% accuracy rate of Autopsy participants. As this task required that the user open a zip file from the original dataset, this may partially explain the lowered accuracy rating of

Insight, as Autopsy simplifies access to the original source image. It was also identified that a common error made in this task was that participants a zip file in the user's My Documents folder as being the answer to the question. No attempt had been made by the suspect to hide this file, and as such this was not the correct answer. The implications of this lowered accuracy rate will be discussed in the next chapter.

When these accuracy results are analysed statistically using an N-1 Chi-Squared test, it can be seen that of the 5 tasks, there is no statistically significant difference between accuracy rates of the Autopsy 3 and Insight tools when using a significance value of 0.05. This can be interpreted positively, as it demonstrates that in the majority of cases, the Insight tool leads to accuracy rates which are similar to the control tool. However, in the case of Question 5, we can see that the p-value was calculated as 0.0004598. This indicates that the difference in accuracy rates between the two tools for this task can be considered as statistically significant. In this case, the accuracy rate of Insight participants was significantly greater. This metric allows us to reject the null hypothesis in relation to this task, as it shows a clear increase in accuracy.

6.2.3 Likert Feedback Results

After completing each task, each user was asked to rate how easily they were able to complete the task with their assigned tool. This took the format of a Likert-style scale which asked the participant: "On a scale of 1 to 5, how easy was it to complete this task with the assigned tool?" On this scale, 1 represented "Very Difficult, 3 represented "Neutral" and 5 represented "Very Easy. Users were asked to provide this feedback so that it could be assessed whether the user groups showed a particular preference for one tool over another.

This information makes it possible to correlate any user feedback where there are a high number of “Very Difficult” or “Difficult” ratings, with reduced accuracy rates. This can also show that the participants were aware of their struggle to find the correct information. Conversely, if the feedback shows that there were a high number of “Easy” or “Very Easy” ratings, but with a low accuracy rating for the task, this can potentially indicate that the participants were of the impression that they had easily found the correct answer, but in fact were incorrect. Instances of this may be interpreted as a cause for concern, as it could indicate that the information presented by the application may not be clear to the end user.

The results of the Likert scale feedback participants can be seen in Figure 6.2. For simplicity, when discussing these results, “Very Easy” and “Easy” ratings will be referred to as “positive”. Conversely, “Difficult” and “Very Difficult” feedback will be referred to as “negative”. A full breakdown of Likert results can be found in Appendix C. The results from this facet of the user study were also subjected to a Mann-Whitney U test (Table 6.3) to determine whether the difference in responses for both tools could be considered to be statistically significant when a significance value of 0.05 is utilised.

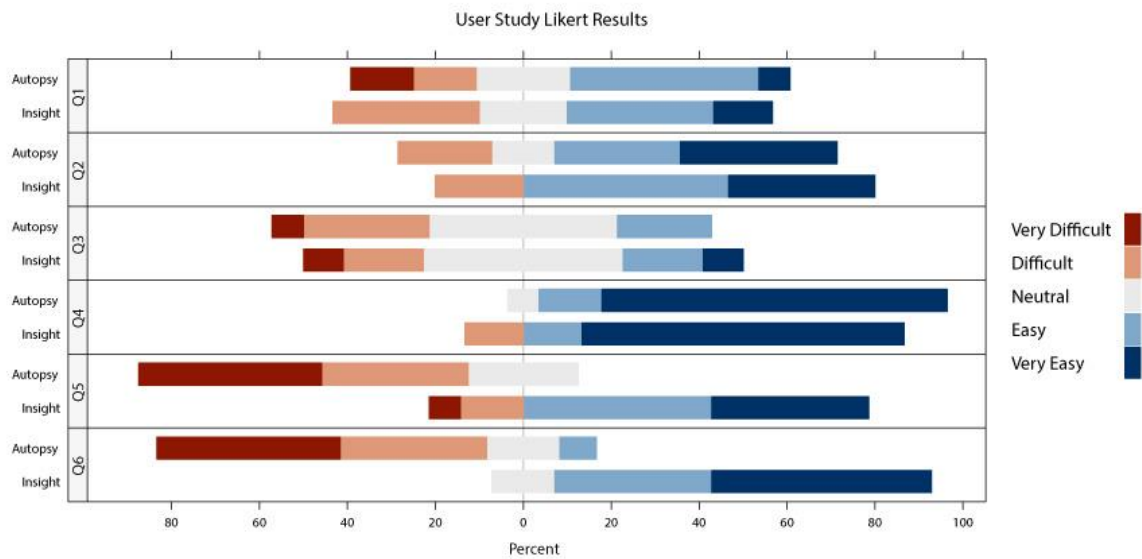


Figure 6.2 - Likert-style feedback ratings for each question

	Q1	Q2	Q3	Q4	Q5	Q6
Mann-Whitney U Value	101.5	99	68.5	112.5	19.5	6.5
Mann-Whitney p-Value ($\alpha = 0.05$)	0.8915	0.7988	0.6432	0.6834	0.0007602	0.00005073

Table 6.3 - Mann-Whitney test results

Although the feedback from users does not serve the purpose of validating or disproving the hypothesis, it is important to interrogate these results as they provide valuable insight into the attitudes of the participants towards each of the tools. A tool which is statistically more efficient to use when based on investigation time, but which users find consistently difficult to use cannot truly be considered to be useful. If this was the case and user's had a more user-

friendly, yet less efficient tool available to them, they will often favour usability over efficiency.

From the 6 tasks presented to the user, we can see from Figure 6.2 that in 4 of these tasks, participants responded more positively to Insight in comparison to Autopsy. In Questions 1 and 4, the participants responded more favourably to Autopsy. However, when looking at both of these Questions, it can be seen from these responses that the preference for Autopsy was not overwhelmingly large. The responses to Q1 show a slightly higher proportion of positive responses for Autopsy, however, Autopsy also received a number of “Very Difficult” ratings. Insight on the other hand, received no “Very Difficult” ratings, and a higher proportion of “Very Easy” responses.

Focus should be drawn to the feedback responses for Questions 5 and 6. Question 5 asked the participant to identify which university the suspect had visited. This task required the user to explore the full dataset in search of evidence. Not only was the accuracy rating of this task for the Insight participant group substantially higher and statistically significant, with Insight participants scoring 80% and Autopsy 3 only scoring 9.09%; the feedback from participants was also substantially more favourable for Insight. It can be seen from Figure 6.2 that the large majority of Insight participants responded positively, whereas there were no positive responses from Autopsy 3 participants. When the feedback responses for this task are subjected to analysis using a Mann-Whitney U test, the calculated p-value is 0.0007602. This indicates that the preference for the Insight tool is statistically significant. As the format of Question 5 involves corroborating a fragment of external information with patterns of user behaviour, based on the results given in this experiment, it can be concluded that the Insight

tool provides a much more intuitive way for the user to explore these narratives of user behaviour, to allow them to discover corroborative evidence. In visualising bursts of web browsing activity, the participants are able to check the beginning of these browsing sessions for any indications that the suspect is connected to a foreign network, i.e. captive portal web pages. Participants from the Autopsy group clearly showed a difficulty in answering this question correctly, as reflected in the accuracy rates. They were also aware of their inability to answer the question, as reflected by the negative trend of user feedback.

In regards to Question 6, although a statistically significant increase in accuracy rates was not demonstrated, a significantly greater trend of positive user feedback for the Insight tool was shown, with a p-value of 0.00005073. The format of Question 6 provides the user with a timestamp and asks them to match this to a suspicious event which had occurred on the suspect device. Although both participant groups demonstrated similar accuracy rates, it can be seen from the user feedback that the user feedback from the two groups was significantly different. The Insight group provided mostly positive feedback for this task, with a small percentage providing neutral feedback. There were no negative responses for the Insight tool for this task. Conversely, the feedback provided from Autopsy user can be seen to be mostly negative, with a small number of neutral and positive responses. From these results, it can be concluded that although both groups were equally able to complete the task correctly, the task was substantially easier to complete with the Insight tool in comparison to the Autopsy 3 tool.

6.2.4 User Feedback

Once each participant had completed all of the main 6 tasks, they were asked to click the “End Experiment” button on the timer application so that their time would be recorded for

only how long it took them to complete these tasks. The next question on the task sheets was to gather qualitative feedback from the participants. This section asked for their opinions on the application they were using in general, whether there were any features of the application which they found difficult or frustrating to use, and whether there were any features of the application which they found very useful. They were also given an opportunity to provide any other comments which they felt were relevant. Some simple thematic analysis was conducted on these points of feedback, to identify key advantages and disadvantages of each application as identified by the participants. The anonymous nature of each participant submission served to give the opportunity to each participant to express the opinions of the application they were using honestly.

In this thematic analysis, all of the feedback from users was read and categorised to allow common themes to be identified. These themes were further categorised as positive, neutral or negative, thus allowing the major themes from each to be identified. Although the feedback task was split into a number of different questions, for the purpose of this analysis, all feedback has been analysed as a whole. The reason for this is that some participants may have mentioned a feature they did not find useful in one section rather than another. For example, a participant may have answered the section regarding features they found difficult or frustrating in a vague way, but may have then identified a feature they found difficult in the section for any other comments. As such, treating all feedback as a whole is the preferred method. It should also be noted that when participant feedback was analysed, it was identified that a significant number of participants from both user groups had identified the applications as crashing occasionally. Although this is valuable feedback, for the purposes of this research, it shall not be considered or included in any analysis. The reason for this is that neither Insight nor Autopsy 3 are commercial grade tools. Insight is a proof-of-concept

prototype tool, and Autopsy 3 is a free and open-source piece of software developed by a community of volunteers. Although it is important for a tool being used for large scale investigative work to be stable, any significant analysis of the stability of the tools involved in this research would be redundant.

The tables which can be found in Appendix D outline the most common positive and negative feedback themes occurring in each of the participant groups.

The thematic analysis of the participant feedback shows that comments regarding the user interface of the Insight application are generally positive. Participants noted that the user interface was clear and user friendly, with number commenting positively on the visual nature of the application, including the use of colour coding to identify different event types. It was often mentioned that it took the participants a short time to familiarise themselves with the application before they felt comfortable using it. Many of the participants also commented positively on the search and filtering functions of the application, stating that it worked well and made finding the required information straightforward. The negative aspects of the Insight application which were identified by participants was the difficulty faced in identifying individual events when there are a large number of events which all occurred at the same time. This is usually due to an automated process creating a number of events all at the same point in time, which can easily overwhelm the ability of the timeline to display all of the events in a useful manner, even when the zooming function is utilised. Some of the participants suggested it would be useful if a vertical scrolling function was provided so that these events could be spaced out further. A number of people also recognised a bug in the timeline scrolling feature which meant that the timeline would suddenly jump back to the

start of the dataset. This bug unfortunately was not recognised in the pilot user study, and as such was not fixed before the main experiment took place. However, from the feedback we can see that the participants considered it a frustration, but not something that prevented them from using the application.

Feedback provided by participants who were using the Autopsy 3 application was noticeably divided. A number of the participants praised the user interface of the application, stating that it was “straight-forward and intuitive” and also “well-organised”. They also commented on the ability of Autopsy 3 to integrate with the filesystem of the suspect device, stating that this made it easy to search through files. It was also included in the feedback section in which the participants commented on what features of the application they found to be most useful. Conversely, there were a significant number of participants who were highly critical of the Autopsy 3 application. A number of the participants felt that the user interface was unintuitive and some even described it as “clunky”, “awkward” and even in one case it was described as “terrible and confusing”. A number of participants also felt that the search functionality of the Autopsy 3 application could be improved, and that as dates could not be searched, this was problematic.

It is interesting to note that, overall, the Autopsy 3 application seemed to receive a larger volume of highly critical feedback from participants than the Insight application, especially when considering the overall usability of the applications. Although there were some instances of negative feedback given for the Insight application, it was generally limited to frustrations with certain features and a number of bugs in the application. One participant did

provide highly negative feedback; however, it was limited to the word “USELESS” in the general feedback section, with no further explanation.

6.3 Conclusion

In this chapter, the various results from the user experiment were discussed. Three elements of the responses by participants were analysed: the time for each participant to complete the experiment, the Likert-style feedback provided by each participant upon the completion of each task, and the overall qualitative feedback given by the participants at the end of the experiment regarding their overall experience with their allocated application. As discussed, the completion times of the participants were recorded and assessed to determine whether the overall efficiency of the investigation is increased when using the Insight tool. As can be seen from the results for this element of the experiment, there was a 14.02% reduction in investigation times when comparing the mean completion time of both participant groups. However, when subjected to t-test analysis, it can be seen that there is no statistically significant difference between the completion times for these groups. As such, we cannot reject the null hypothesis based on these results.

The accuracy rates of the participants were also analysed. The intention of this metric was to ensure that as a minimum, the Insight application would demonstrate accuracy levels which were comparable to Autopsy 3. This would show that at very least, the Insight application allowed the end user to find comparable volumes of correct evidence as a mature digital forensics tool. This metric was also used to test if the null hypothesis could be rejected, on the basis that increased accuracy levels would be perceived as an important benefit. In the majority of the tasks given to the participants, the accuracy rates were slightly higher for the

Insight tool, although when subjected to analysis with an N-1 Chi Squared test, it was shown that in 5 of the 6 tasks, the difference in accuracy rates could not be described as statistically significant. However, the accuracy rate results for Question 5 are demonstrated to be much higher for the Insight tool, with the Autopsy 3 tool only reaching an accuracy rating of 9.09%, compared with the 80% accuracy rate achieved by Insight participants. This result is shown to be statistically significant with a p-value of 0.0004598. This result allows the null hypothesis to be rejected, as we can see there is a clear accuracy benefit provided in certain scenarios when an exploratory information visualisation technique is applied to the investigation.

Furthermore, the feedback of the participants for each task in the format of Likert-style data was analysed to draw conclusions about each participant's experience after completing each task. From these results, it was shown that for the first 4 tasks, there was no statistically significant difference between the participant feedbacks when analysed using a Mann-Whitney U test. The user feedback for Questions 5 and 6, however, showed a strong preference for the Insight application. When these results are combined with the thematic analysis of written user feedback, it can be concluded that in general there is a stronger preference for the Insight application in general.

These results will be discussed in-depth in the next chapter in regards to the overall outcome of this research, future work, and how the results from this research validate the research hypothesis.

Chapter 7 - Discussion

7.1 Introduction

In this chapter, the results which were gathered and analysed in the previous chapter will be discussed in further detail. This discussion will focus on what these results mean in regards to providing an answer to the research question - does the utilisation of exploratory information visualisation techniques in digital forensic investigations provide benefits in regards to efficiency or accuracy?

The discussion will look at the time to completion results and accuracy rating results from the participant study. These results will be discussed with a view to examining what they show, and how they provide an answer to the research question. The feedback provided by the participants, both in the form of Likert-style ratings at the end of each task, and as written comments at the end of the experiment, will be discussed and conclusions drawn about the potential relationships between this feedback and the result data. This participant feedback will also be used to discuss potential issues with each of the two test applications, along with potential resolutions for these issues, and how they could lead to potentials for further work.

7.2 Discussion of Results

7.2.1 Completion Time

One of the metrics used in this research was the time taken by each participant to complete the experiment. Often in industries which utilise digital forensics investigations, such as law enforcement, there is too great a volume of data to be processed in a timely manner (Casey, Ferraro and Nguyen, 2009). In many cases, this leads to a large backlog of cases waiting to be

processed. For this reason, the completion time was selected as one of the core metrics against which the hypothesis would be tested. Ideally, a tool would provide a way in which investigators could analyse a suspect device in a significantly shorter time than is currently possible with traditional tools.

As was identified in the literature review, the tools which are currently available to the digital forensics industry are largely text-based. It was hypothesised that by applying information visualisation techniques to the datasets, the investigator may be able to discover narratives of behaviour and corroborative evidence more effectively. This metric was tested by conducting a user experiment on a group of 29 4th year undergraduate and Masters postgraduate digital forensics students who were all studying the same module. 14 of the participants were instructed to use the Autopsy 3 tool. These participants were designated as the control group for this experiment. The remaining 15 participants used the Insight application to complete the experiment. Both groups were timed, and then the timings from both groups were analysed to ascertain whether there was any significant reduction in completion time in the Insight group when compared to the Autopsy test group.

As was shown in the last chapter, the results given do not allow us to reject the null hypothesis: ‘the utilisation of exploratory information visualisation techniques in digital forensics investigations provides no benefit in regards to effectiveness. When the mean value of the two test groups was analysed, a 14% reduction in completion time for the Insight test group in comparison to the Autopsy control group can be seen. However, when the two time samples were analysed using a two-tailed unpaired t-test, the resulting p-value of 0.4382 does not allow us to reject the null hypothesis. Therefore, based on this metric, we cannot say that

there is any efficiency gain when using an application which utilises information visualisation when compared to a traditional text-based tool. It would, however, be unwise to rule out the possibility altogether that an efficiency gain when using these tools is possible. The results collected by this experiment do not statistically support the hypothesis, however, it is possible that there may be some scenarios in which the tool would be especially suited and would provide substantial efficiency gains.

The experiment which was conducted provided the participants with a variety of tasks which involved looking at different data types, and utilising the applications in different ways. It would be useful to analyse completion times for each type of task in order to assess whether there are any substantial efficiency gains when the user is presented with a certain scenario. In order to analyse this, it would be necessary to present the user with only one task and to record a completion time for this. Due to the restrictions faced during this research with regards to the limited number of datasets available, and also on the number of suitably experienced participants, it was not possible to take this approach. Doing so would have given a very small sample size for each question, and would lead to all results being unusable as scientifically valid data.

It could be argued that the approach of timing each question individually for each participant could have been taken in order to give a clearer view on how the overall completion time was distributed across each task. This approach was not taken as it was argued that not only would it require that the participants interact with the timer application after every task, but it would also assume that each task was entirely self-contained. This is clearly not the case, as the participant is presented with the entire dataset when starting the application. They may

then wish to sit and browse through the dataset before answering the first question. Such an approach would lead to an unusually high time to completion for the first question, and then unusually low completion times for subsequent tasks. Even if the participant chooses not to browse the entire dataset at the beginning of the experiment; it is inevitable that knowledge from one task would leak into another. For example, if the participant happens to find the answer to a later question, they are likely to be able to locate it quickly again when the need arises, thus leading to synthetically shortened times for some tasks.

As discussed, had a significantly larger pool of participants been available, it would have been useful to examine efficiency on a more granular level. However, for the purposes of this research, this was not feasible. The completion time results which were gained from the experiment in this instance do not support the research hypothesis. It is important, however, to examine other important metrics such as the accuracy of the participant's conclusions.

7.2.2 Accuracy Rate

The accuracy rate of each participant was also recorded when conducting the experiment. This metric was assessed based on whether the participant had correctly answered all sections of a task. Only if no errors were present in the participant's response would the task be considered to have been completed accurately. If errors were present in any section of the task, then the entire task for that user would be marked as incorrect. The purpose of collecting this data was to allow the research hypothesis to be compared against a second metric. Although the completion time data does not support the research hypothesis, as no statistically significant difference from the control group could be observed, this does not preclude the possibility of benefits being afforded in the accuracy of investigations. Before

discussing the accuracy rating results, it is important to examine how the previous metric of completion time relates to the accuracy metric. It can be seen from statistical analysis of the completion time data that there is no significant difference between the completion times of Insight participants and the control group. Although this means that Insight does not significantly shorten investigation time, it also means that there is no significant disadvantage to using the Insight application. This data reflects the ability of Insight to allow a user to complete an investigation in a timescale which is comparable to that of an industry-standard tool. This outcome allows us to move on and assess the second metric, accuracy rate, in order to validate the research hypothesis. Had the completion time results shown a significant increase in investigation time when using the Insight tool, this would have effectively nullified any benefits seen in other areas.

When looking at the accuracy data derived from the experiment, as discussed in the previous chapter, it can be seen in tasks 1-4 and 6 that when subjected to analysis by means of an N-1 Chi Squared test, there is no statistically significant difference in accuracy rates between both applications. This is a positive result in its own right, as it indicates that in these tasks, both groups were able to complete the tasks to a similar level of accuracy. This validates that the information available to the user when using the Autopsy 3 tool, is also clearly available when using the Insight tool. This also validates that the Insight tool has been developed correctly, as the dataset used by Insight is derived entirely from the processed output of Autopsy 3, as was discussed in the Methodology chapter.

From the analysis of the accuracy data, it can, however, also be seen that there is a marked difference in the accuracy rates achieved by participants in both test groups for Question 5.

For this task, the participants were told that the suspect had visited a university at a point in time while using the device. They were asked to give details of which university was visited, the date on which it was visited, and to provide evidence of how they reached these conclusions. Of the Insight test group, 80% of participants answered all sections of this task correctly. However, of the Autopsy 3 test group, only 9.09% of participants managed to answer all sections of this task correctly. When subjected to analysis via N-1 Chi Squared testing, this equates to a p-value of 0.0004598. As a significance value of 0.05 has been used for all results in this research, this indicates that in this instance, the result can be considered to be statistically significant. As this question required the user to explore the majority of the dataset themselves, with very little information to guide to them to where they should look, this is further reinforced when the user feedback is taken into account. Many of the participants in the Autopsy 3 user group actively commented on how difficult they found Autopsy 3 to use when completing the final two tasks, which included Question 5.

It can also be assured that the information was available to the participants in the Autopsy 3 participant group, as all information shown on the Insight visualisation was taken directly from the output of Autopsy, as was explained in the Methodology. The similar accuracy rates between the applications when looking at the other five tasks also reinforces this point. Thus, it can be concluded that the presentation of the information by the Insight tool was statistically superior to this. In conclusion, the accuracy results taken from Question 5 allow us to reject the null hypothesis - 'the utilisation of exploratory information visualisation techniques in digital forensics investigations provides no benefit in regards to effectiveness'. These results demonstrate that in certain scenarios, the presentation of digital forensics information in the format of a visual model can increase the ability of the investigator to accurately find required information in a suspect dataset.

7.2.3 Likert Scale Feedback

As was discussed in the Results chapter, the participants were polled for their feedback on how easily they found each task to complete with the tool they were allocated. This scale took the format of a scale of 1-5, with a rating of 1 representing 'Very Difficult', a rating of 3 representing 'Neutral', and 5 representing 'Very Easy'. The purpose of collecting this information was to reinforce any conclusions drawn from the efficiency or accuracy result sets and to allow new conclusions to be drawn regarding user experiences with each of the two test tools.

Once these results had been collected, they were subjected to statistical analysis by means of a Mann-Whitney U test in order to determine if the participants in both groups had statistically significant differences in experience when trying to complete each task. In regards to the first four of the tasks completed by the participants, there was no statistically significant difference found between the experiences of both test groups. When looking at the results for the final two tasks, however, the participants reported consistently opposing experiences. The p-values for both of these tasks when subjected to a Mann-Whitney U test show that the difference in user feedback between the Autopsy and Insight participant groups is statistically significant. The feedback from participants shows that for Questions 5 and 6 is significantly more positive in the Insight participant than in the Autopsy 3 group. Feedback for both of these tasks was almost entirely positive for Insight and almost entirely negative when looking at the Autopsy group. As there were 2 instances of a greater positive feedback for the Insight tool, and 4 instances of statistically insignificant difference in feedback for either tool, it can be concluded from these results, that the Insight tool provides a more

intuitive and useful user experience in comparison to the Autopsy 3 tool. This conclusion is reinforced when the written feedback of the participants is analysed, as discussed in the previous chapter. It can be seen that there is a general trend towards positive feedback towards the Insight tool. Negative criticisms for this tool generally involve bugs which exist in the prototype version, and difficulty in viewing individual events where there were a large number that had occurred all at the same point in time. Feedback for the Autopsy tool, however, was very conflicted, with a number of participants praising its ease of use and well-organised user interface. Other participants, however, provided feedback which was highly critical of the application and its features, stating that many some features such as the search function were not entirely useful and that the application in general was not useful, particularly in relation to its user interface, which was viewed by some as being clunky and too simplistic.

7.3 Research Outcomes

7.3.1 Validation of Research Hypothesis

This research has involved a number of different metrics which have been collected and assessed; efficiency, accuracy and user experience. It is important to consider these metrics, and the relationships between them. Analysing them in isolation to the others allows us to draw some conclusions, but does not examine how each of these metrics impacts upon the others. Although analysing these metrics in isolation would technically allow us to either reject or accept the research hypothesis, it would be unwise to do so. For example, hypothetically speaking, had this research had shown a statistically significant decrease in investigation time when using the Insight tool, this would have allowed us to reject the null hypothesis, as a clear benefit in regard to a significant increase in effectiveness would have been shown. However, in this scenario, the research could have then gone on to identify a significant decrease in the accuracy levels for Insight participants. Examining the metrics in isolation would allow the research hypothesis to be validated, as benefits had been demonstrated in regards to efficiency, but this would entirely ignore the fact that the increase in efficiency was paired with a marked decrease in accuracy. It is common sense to realise, in this hypothetical scenario, that this trade-off is entirely unacceptable and leads to a tool which is unfit for use in any investigation. As such, the outcomes of this research were identified by considering all metrics as a whole, and with consideration as to their relationships.

As can be seen by the results regarding efficiency of investigations between the two test applications, there was no statistically significant difference between the two test groups. Although this does not allow the null hypothesis to be rejected at this stage, if the Autopsy 3 application is taken as a benchmark of an industry-standard tool, then these results show that

the Insight application allows an investigation to be completed in a timeframe which is comparable to current industry-standard tools. This allows the other benefits which may be afforded by the utilisation of information visualisation techniques to be explored, with the knowledge that they will not be at the cost of an increased investigation time.

The accuracy metric shows that in one of the tasks, there was a substantial increase in accuracy rates when using the Insight tool. As was discussed, this allows us to reject the null hypothesis as it indicates an example of a significant benefit when implementing information visualisation techniques. Combined with the knowledge that this benefit exists, without sacrificing overall investigative efficiency, we can accept the research hypothesis - 'the utilisation of exploratory information visualisation techniques in digital forensics investigations provides benefits in regards to effectiveness. This conclusion is further reinforced when the Likert scale participant results are brought into consideration. Not only is a significant increase in accuracy demonstrated in one instance, but the participant feedback for the same task, Question 5, also shows that participants generally found it easy to find the required information when using the Insight tool, whereas the large majority of Autopsy participants reported that they encountered difficulty in completing the task. This result is interesting as this question presented the user with a fragment of external knowledge, i.e. that the suspect visited a university, and required them to explore the dataset to find corroborative evidence of this fact. This suggests that Insight may be much more of an effective tool when the user can target their searches based on pre-existing knowledge. The Likert feedback results for Question 6 also shows a similar outcome. In this task the participant was given a point in time, and was required to find an event which occurred at this time showing suspicious user activity. Although the accuracy rate for this task was not significantly different between user groups, the feedback which is significantly more favourable for the

Insight tool. This validates the use of a timeline style visualisation format, as the feedback shows that when dealing with chronological information, the Insight tool made this much simpler. Written feedback from the Autopsy participants also showed a number of cases where the search feature was criticised as not allowing the user to search by time, and as such, they found this task generally difficult to complete.

To summarise, this research allows the research hypothesis to be accepted, as there is evidence to show that by applying information visualisation techniques to digital forensics investigations, there exists a potential for significant improvement in accuracy rates while maintaining an efficiency rate which is comparable to industry-standard tools. Also, there is evidence to suggest that the overall user experience when utilising tools which integrate information visualisation techniques is more positive than when traditional text-based tools are utilised.

7.3.2 Research Significance

The aim of this research was to determine whether the application of information visualisation to the analysis phase of a digital forensic investigation yields any significant benefit to the end user in terms of efficiency or accuracy gains. This was to be achieved through the use of a number of case studies to develop a framework and appropriate visual model which would lead to the development of a prototype application. This prototype application would then be tested on a suitable user group to assess any potential benefits afforded by the use of information visualisation in this field.

As can be seen from the results, the original aims of this research have been satisfied, and results from the research suggest that the application of information visualisation to digital forensic investigations, specifically the analysis phase, does in fact provide benefits in the form of increased accuracy rates. Not only this, but this research has found that the user experience when using tools which utilise information visualisation techniques is generally more positive, when compared with the user experience when using a traditional text based tool. This not only applies to the Autopsy 3 tool which was used as the experiment control tool; one Insight participant also noted that the tool provided them with much more information when compared to another text-based digital forensics tool, Zeitline.

These results provide the groundwork and motivation for further research and development in this area. Also, as a prototype tool, the Insight application itself provides a framework on which further work could be based on. The modular nature of the application provides a number of way in which sections of the application could be integrated into future research. For example, the pre-processor section of the application could easily be taken and integrated into another project, providing a standardised way to extract information from the Autopsy database.

7.3.3 Research Limitations

It should be recognised that there are limitations to this research which limits the number and granularity of the outcomes. The most significant of these limitations is the volume of participants available to take part in any experiments which were conducted. As investigation of a digital forensics dataset requires a base level of knowledge, it was not possible to simply advertise for participants to take part. Instead, it was necessary to identify a group of

participants who all had similar levels of experience, of which there were a limited number available. This was not only necessary so that the participants had that correct knowledge to be able to complete the experiment, but also so that a reliable set of results could be obtained. Any significant variations in participant experience would have the potential to produce inaccurate results.

Another limitation which is common to the testing of all digital forensics tools, is the availability of suitable datasets. In the case of this research, one dataset was available which could be used in an experiment. The “John Doe” case, as has been previously discussed, is a synthetic test case created by Dr Ian Ferguson to assist in the teaching of digital forensics modules, and is usually presented to the students as a final coursework task. This dataset was ideal for experimental use, as it is small enough to allow the participants to complete an experiment in a reasonable length of time. Presenting the participants with a realistic scale dataset would substantially increase the length of time it took the participants to browse that dataset, and this increase would likely lead to difficulties in recording timings, as participants would have to take breaks. It would also likely lead to a number of participants withdrawing from the experiment as they would not be willing to dedicate the required time to it. With an already restricted pool of available participants, this would have been problematic. Readily available realistic scale datasets are also commonly taken from a live system, and as such are likely to contain varying levels of personally identifiable information. The task of ensuring that all of this personal information was found and redacted would not be feasible, and as such this precludes the use of realistic datasets.

7.4 Further Work

This novel aspect of this research was to establish whether information visualisation techniques could be applied to digital forensics datasets in a way which would give the investigator an overall view of the dataset, rather than only a subset of data such as web browsing history. This was carried out through the creation of a new tool which presents the majority of information from a device dataset to the investigator in a visual timeline format. As part of the research, the effect this visual tool had on the effectiveness of digital forensics investigation was evaluated using a group of suitably trained students, tasked with analysing a synthetic test case.

This research demonstrated a number of positive effects on investigative effectiveness such as instances of increased accuracy, and the feedback from participants indicating that the visual Insight tool made reaching a conclusion easier in certain tasks. However, there was no noticeable benefit in respect to the length of time taken by participants to complete the investigation. This is not entirely negative as this shows that the positive aspects of the visualisation are not offset by an increase in investigation time.

However, as the number of available participants was a limitation of the research, it would be beneficial if a similar experiment could be conducted with a larger number of participants, and with a larger test dataset. The reason for this being that this would likely give a more accurate portrayal of whether there are likely to be any efficiency gains when a visualisation tool is applied to investigations of a more realistic scale. It may be the case that the as the user familiarises themselves with the tool, that they become more efficient in using it. This could also be the case with the textual tool however. Ideally, it would be useful to have a number of investigators look at a real case, over a longer period of time using either the Insight or Autopsy tool. The difficulty in this is the acquisition of a real dataset which in most cases would be restricted to law enforcement personnel. If law enforcement personnel were to

be the test subjects involved, it would also be difficult for them to dedicate resources to 2 groups of staff looking at the same case. The research has discussed the backlogs faced by these individuals, and so a “chicken and egg” problem arises. However, involving such individuals would be an almost ideal scenario.

It was recognised after the experiment was conducted that it would have been useful to analyse the time taken to complete each individual task, and to categorise these tasks clearly, based on the type of information the participant is being asked to examine. In doing so, it could give a granular insight into which type of problems both the visualisation based and traditional tools excel at solving. Such conclusions can be extrapolated from the user feedback, however, it would be useful to have statistical time data to provide further insight. The difficulty in doing this in this instance was that each participant would have had to record their times themselves, which lead to a higher risk of errors. And due to the limited time with the participants and a single dataset, this experiment was “one-shot” with no opportunity for repetition. A significantly larger participant pool would have allowed more in-depth experimentation.

When analysing feedback provided by the participants, it was identified that in some cases participants had indicated that they had difficulty in interacting with the suspect dataset. This was specifically related to Question 3 in which the participant was asked to find a ZIP file that the suspect had attempted to hide. They were required to use the tool provided to find information about this file. They could find this information by looking at the Extension Mismatch data provided in the tool which would then direct them to the ZIP file of which the suspect had changed the file extension to .dll. As part of this task, the participant was then

asked to open the file (an appropriate third party tool was pre-installed on the VM) and to provide the names of some of the files contained within. Autopsy users showed to have less difficulty with this task than the Insight users, possibly due to Autopsy's close integration with the source disk image. There are a number of approaches which could be taken to resolve this issue; firstly, the ability to directly access the source disk image could be built into the Insight application; alternatively, the Insight application could be modified so as to work as a plugin to the Autopsy 3 application itself. Autopsy 3 has a number of strengths, and by providing the core timeline functionality of Insight from within the Autopsy 3 application itself, this would effectively combine the strengths of both applications.

When examining the various case studies which contributed to the Insight framework, the approach of 3D hyperbolic visualisation was noted as a novel way to display filesystem information. Although this visualisation format was not suitable for the purpose of this research, it is not without its merits. As the timeline format of Insight is not well suited to displaying file system structure information which is hierarchical rather than chronological; it may be possible for a piece of future work to attempt to integrate the hyperbolic visualisation approach demonstrated in the Walrus application into the main Insight application. Such an approach may provide the user with both visualisations on screen at the same time, and as such would provide them with both chronological event information and filesystem information at the same time. This may have the effect of increasing the accuracy of tasks which involve the examination of primarily filesystem based information. Based on the accuracy ratings from the experiment which was conducted, this was the area in which Insight received a lower accuracy rating than Autopsy.

In summary, recommendations for future work would include expansion of the prototype tool to allow deeper integration with the source disk image, as opposed to its metadata. Although the exact form this would take is unclear at this time; as discussed, it may be of interest to attempt to integrate a visualisation technique which utilises the 3D hyperbolic model to represent this data. Further to the adaptation of this tool, it would be of use to area of research to examine that type of information that each tool excels in representing to the user. Doing so would provide the potential to develop a tool which integrates a number of different display methodologies, which lend themselves well to all types of data encountered by a digital forensics investigator. As the efficiency metric of this body of research showed no significant gains provided through the use of visualisation, it would be interesting to collect data on the length of time taken by participants to complete individual tasks. Such data would allow for specific refinement of the tool which may eventually lead to significant efficiency gains over the course of a realistic scale digital forensics examination.

This researcher notes that at the time of this research, virtual reality technology was in its relative infancy, with hardware such as the Oculus Rift nearing release to general consumers. Although this body of research found 3D models to be somewhat cumbersome and difficult to draw clear conclusions from, it may be an area which has merit in re-revisiting in future. Virtual reality provides a way for the end user to truly immerse themselves in an environment. It is speculated that this may provide new way to display digital forensics visualisations to the end user, in a way which allows them more space to explore the data than ever before, and may be an area of potential future research projects.

7.5 Conclusion

From the analysis of this research task as a whole, it can be seen that there is merit in applying information visualisation techniques to digital forensics datasets in order to support the Analysis phase of an investigation. In displaying event information derived from the suspect device image as an interactive timeline, it has been shown that in certain cases this can lead to significantly increased accuracy rates. This research has provided a framework for future work, as was discussed, and in validating the research hypothesis, this research provides a valid motivation to pursue further research projects in this field.

Chapter 8 - Conclusion

Digital forensics is an increasingly important industry in our society for number of reasons and the number of devices owned by a person and which are connected to the internet is continuously increasing. As valuable sources of evidence, it is vitally important that these devices can be analysed by digital forensics experts in a timely manner, often in order to support law enforcement efforts. However, as was discussed in the literature review chapter, there is a lag in the ability of modern tools to keep pace with the ever increasing number of device. Arguably the least well supported phase of the investigative process is the analysis phase.

This research aimed to examine if the analysis phase of the investigatory process could be better supported through the application of information visualisation techniques to display the dataset to the investigator in a format that would aid the analysis process. This was carried out by examining a number of case studies in which various prototype tools were created, utilising various types of visualisation and technologies. From each of these case studies, the lessons learned from each were fed into the next, to build a gradually evolving framework which allowed the output from the Autopsy 3 pre-processor application to be accessed and visualised by the prototype application. Lessons learned from the case studies included the need for an index file, to accelerate application load times by eliminating the need for the application to repeatedly re-read information from the main Autopsy database; and various technologies were used until the appropriate platform was found to be a desktop application built upon the Microsoft .NET framework, due to the performance issues encountered with other frameworks such as JavaScript based applications.

The prototype application was built with the outcomes of the previous case studies, and developed using a modular approach which allowed functionality of the application to be changed with relative ease. For example, all pre-processing functionality which generates an XML index file from the Autopsy database was developed as a separate application so that if needed, it could be replaced or changed without affecting the main application. So long as the format of the output XML file did not change, the main Insight application would continue to operate. This affords the opportunity for the source of the data to be changed to something other than Autopsy should the need arise.

Once the prototype Insight application had been developed, and tested by a small pilot group with the intention of identifying and resolving any obvious bugs, an experiment was designed for a larger group of participants with a relative level of experience. The experiment aimed to look at whether the prototype tool which had been developed would provide any significant benefit to the end user in regards to the ability to complete the investigation in a shorter time, the ability to draw conclusions about the data more accurately, and whether the user experience when using this application was favourable.

The results from this experiment, as addressed in the Discussion chapter, show there is evidence that when the investigator is given a tool to use which utilises information visualisation techniques, in this case in the format of a timeline, there is a potential for significant increase in accuracy in some scenarios. This is paired with a generally more favourable user experience when compared to traditional text-based tools. The investigation duration also is not significantly different to when a traditional text-based tool is used, meaning that there are no substantial disadvantages to utilising information visualisation

techniques in the analysis phase of the investigation. This experiment allows the research hypothesis - ‘the utilisation of exploratory information visualisation techniques in digital forensics investigations provides benefits in regards to effectiveness’ - to be accepted.

In conclusion, this research has provided evidence that the incorporation of information visualisation techniques into tools designed to support the analysis phase of a digital forensic investigation has merit. It is suggested that further work be carried out in this area in order to further develop and refine these techniques, and to test their use in large scale investigation, so as to gain insight into further benefits which may be afforded. It is further hypothesised that if applied to a large scale investigation, the statistically insignificant efficiency gains which were suggested by this research, may be amplified and may result in more substantial efficiency improvements. Even if this is found not to be the case, the accuracy gains, and the overall more positive user experience from tools of a visual format are worth pursuing.

The research has also contributed a framework, and a set of outcomes on which further work can be based. The outcomes from the case studies examined in the Methodology provide a set of recommendations and points to consider when developing an information visualisation format which links to a digital forensics dataset. It may also be possible to develop these outcomes further and to conduct further similar case studies, with a view to creating a formalised set of ‘best practices’ for developing visualisation tools for digital forensic datasets.

References

Agarwal, A. et al., 2011. *Systematic digital forensic investigation model*. International Journal of Computer Science and Security (IJCSS), 5(1), pp.118–131.

Apache Storm. 2016. *Apache Storm*. [Online] Available at: <http://storm.apache.org/>

Avg.com. 2015. *AVG | The History of the Internet*. [Online] Available at: <http://www.avg.com/history-of-internet> [Accessed 31 May 2015].

Azzam, T. et al., 2013. *Data visualization and evaluation*. New Directions for Evaluation, 2013(139), pp.7–32.

Bishop, I.D. & Stock, C., 2010. *Using collaborative virtual environments to plan wind energy installations*. Renewable Energy, 35(10), pp.2348–2355. Available at: <http://dx.doi.org/10.1016/j.renene.2010.04.003>.

Bostock, M. 2016. *D3.js - Data-Driven Documents*. [Online] Available at: <https://d3js.org/>.

Carrier, B., 2009. *The Sleuth Kit and Autopsy: forensics tools for Linux and other Unixes*, 2005. Available at: <http://www.sleuthkit.org>

Casey, E., 2010. *Digital dust: Evidence in every nook and cranny*. Digital Investigation, 6(3-4), pp.93–94. Available at: <http://dx.doi.org/10.1016/j.diin.2010.02.002>.

Casey, E., Ferraro, M. & Nguyen, L., 2009. *Investigation delayed is justice denied: Proposals for expediting forensic examinations of digital evidence*. Journal of Forensic Sciences, 54(6), pp.1353–1364.

Collins, M., 1999. *Telecommunications crime - Part 2*. Computers and Security, 18(2), pp.683–692.

Conti, G. 2007. *Security data visualization*. San Francisco: No Starch Press.

Dees, T., 2002. *EnCase*. Law & Order, 50(10), pp. 102-103.

FBI, 2006. *Innocent Images International Task Force*. [Online] Available at: http://www.fbi.gov/news/stories/2006/february/innocent_images022406 [Accessed 31 May 2015].

Fekete, J. et al., 2008. *The Value of Information Visualization*. *Information Visualization*, 4950, pp.1–18.

ForensicsWiki. 2015. *Encase image file format - ForensicsWiki*. [Online] Available at: http://forensicswiki.org/wiki/Encase_image_file_format.

Garfinkel, S.L., 2013. *Digital Forensics*. American Scientist, 101(5), pp.370–377.

Heer, J., Card, S.K. & Landay, J.A., 2005. *Prefuse: a toolkit for interactive information visualization*. In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp. 421–430. Available at: <http://dl.acm.org/citation.cfm?id=1055031>

Henry, N., Fekete, J.D. & McGuffin, M.J., 2007. *NodeTrix: A hybrid visualization of social networks*. IEEE Transactions on Visualization and Computer Graphics, 13(6), pp.1302–1309.

Hinrichs, U., Schmidt, H. & Carpendale, S., 2008. *EMDialog: Bringing information visualization into the museum*. IEEE Transactions on Visualization and Computer Graphics, 14, pp.1181–1188.

Hughes, T., Hyun, Y. & Liberles, D.A., 2004. *Visualising very large phylogenetic trees in three dimensional hyperbolic space*. BMC Bioinformatics, 5(1), pp.1–6. Available at: <http://dx.doi.org/10.1186/1471-2105-5-48>.

Isaacs, J.P. et al., 2011. *Immersive and non immersive 3D virtual city: Decision support tool for urban sustainability*. Electronic Journal of Information Technology in Construction, 16(January), pp.151–162.

Kent, K. et al., 2006. *Guide to integrating forensic techniques into incident response*. NIST Special Publication, pp.800–886.

Leschke, T. & Nicholas, C., 2013. *Change-link 2.0: a digital forensic tool for visualizing changes to shadow volume data*. Proceedings of the Tenth Workshop on Visualization for Cyber Security, D, pp.17–24. Available at: <http://dl.acm.org/citation.cfm?id=2517960>.

Liu, Y., Barlowe, S., Feng, Y., Yang, J. and Jiang, M., 2013. Evaluating exploratory visualization systems: A user study on how clustering-based visualization systems support information seeking from large document collections. *Information Visualization*, 12(1), pp. 25-43.

Lowman, S. & Ferguson, R.I. 2011. *Web History Visualization for Forensic Investigations*, Proc. of Cyberforensics 2011, Univ. Strathclyde

Martínez-Cámara, E. et al., 2012. *Sentiment analysis in Twitter*. *Natural Language Engineering*, 20(1), pp.1–28. Available at: http://www.journals.cambridge.org/abstract_S1351324912000332.

McDermott, I.E., 2015. *Ransomware: Tales From the CryptoLocker*. *Online Searcher*, 39(3), pp. 35-37.

Microsoft. 2016. *Compiling to MSIL*. [Online] Available at: [https://msdn.microsoft.com/en-us/library/c5tkafs1\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/c5tkafs1(v=vs.85).aspx)

Mono Project. 2016. *Mono*. [Online] Available at: <http://www.mono-project.com/>

Munzner, T. & Burchard, P., 1995. *Visualizing the structure of the World Wide Web in 3D hyperbolic space*. Proceedings of the first symposium on Virtual reality modelling language - VRML '95, pp.33–38.

Nassi, I. & Shneiderman, B., 1973. *Flowchart techniques for structured programming*. ACM SIGPLAN Notices, 8, pp.12–26.

Nataraj, L. et al., 2011. *Malware images: visualization and automatic classification*. Proceedings of the 8th International Symposium on Visualization for Cyber Security.

Nightingale, F., 1858. *Notes on matters affecting the health, efficiency, and hospital administration of the British Army*. Available at: <https://archive.org/details/b20387118>

Noblett, M., Pollitt, M & Presley, L. 2010. *Recovering and Examining Computer Forensic Evidence*. Forensic Science Communications, FBI, 2(4). Available at: <http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct2000/index.htm/computer.htm>

Parker, D.B., 1976. *Crime by computer*, Scribner.

Playfair, W. 1796. *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*

Pollitt, M.M., 2007. *An ad hoc review of digital forensic models*. In Systematic Approaches to Digital Forensic Engineering, 2007. SADFE 2007. Second International Workshop on. IEEE, pp. 43–54. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4155349

Pollitt, M., 2010. *A History of Digital Forensics*. Kam-Pui Chow; Sujeet Sheno, ed. Advances in Digital Forensics VI. Hong Kong, China: Springer, pp. 3–15.

Quist, D.A. and Liebrock, L.M., 2011. *Reversing Compiled Executables for Malware Analysis via Visualization*. Information Visualization, 10(2), pp. 117-126.

Roussev, V., Quates, C. & Martell, R., 2013. *Real-time digital forensics and triage*. Digital Investigation, 10(2), pp.158–167. Available at: <http://dx.doi.org/10.1016/j.diin.2013.02.001>.

Sankey, H.R., 1898. *Introductory note on the thermal efficiency of steam-engines*. In Minutes of Proceedings of The Institution of Civil Engineers (pp. 278-283).

Sanders, C. & Smith, J., 2014. *Chapter 9 - Signature-Based Detection with Snort and Suricata*, Applied Network Security Monitoring. In Boston: Syngress, pp. 203–254. Available at: <http://www.sciencedirect.com/science/article/pii/B978012417208100009X>.

Schmerl, S. et al., 2010. *Explorative Visualization of Log Data to Support Forensic Analysis and Signature Development*. 2010 Fifth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering, pp.109–118. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5491960> [Accessed October 5, 2011].

Selman, D. 2002, *Java 3D programming*, Manning, Greenwich, Conn.

Shneiderman, B., 1992. *Tree visualization with tree-maps: 2-d space-filling approach*. ACM Transactions on Graphics, 11(1), pp.92–99.

Shneiderman, B., 1996. *The eyes have it: a task by data type taxonomy for information visualizations*. Proceedings 1996 IEEE Symposium on Visual Languages, pp.336–343.

Shneiderman, B., 2008. *Extreme Visualization: Squeezing a Billion Records into a Million Pixels*. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. SIGMOD '08. New York, NY, USA: ACM, pp. 3–12. Available at: <http://doi.acm.org/10.1145/1376616.1376618>.

Snow, J., 1855. *On the Mode of Communication of Cholera*, John Churchill. Available at: http://books.google.co.uk/books?id=-N0_AAAAcAAJ.

Spiceworks. 2016. *Free Help Desk, Network Monitoring, and IT Community*. [Online] Available at: <http://www.spiceworks.com/>.

SQLite, 2016. *SQLite Home Page*. [Online] Available at: <https://www.sqlite.org/>.

Stephens, M., 2012. *floatingsheep: Church or Beer? Americans on Twitter*. [Online] Floatingsheep.org. Available at: <http://www.floatingsheep.org/2012/07/church-or-beer-americans-on-twitter.html>

Stoll, C. 1989. *The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage*. Pocket Books

Tan, S. et al., 2014. *Interpreting the public sentiment variations on Twitter*. IEEE Transactions on Knowledge and Data Engineering, 26(5), pp.1158–1170.

Teerlink, S. & Erbacher, R., 2006. *Improving the computer forensic analysis process through visualization*. Communications of the ACM, 49(2), pp.71–76. Available at: <http://dl.acm.org/citation.cfm?id=1113073> [Accessed January 31, 2012].

ThinkBroadband. 2014. *UK Broadband Factsheet - Q1 2014*. [Online] Available at: <http://www.thinkbroadband.com/factsheet/broadband-factsheet-q1-2014.pdf> [Accessed 31 May 2015].

U.S Government. 1999. *United States vs. Brunette*, 76 F.Supp.2d 30. Available at: http://www.leagle.com/decision/199910676FSupp2d30_1101.xml/U.S.%20v.%20BRUNETTE

Whitcomb, C. 2002. *An historical perspective of digital evidence*, International Journal of Digital Evidence, 1(1).

White, T., 2012. *Hadoop: The definitive guide*. , O'Reilly Media, Inc.

Appendix A – Experiment Participant Sheets

Insight Usability Study - Information Sheet

Introduction

The Insight application was developed to assess whether the application of visualisation techniques can improve the efficiency of digital forensics investigations when compared to traditional text-based tools.

This study aims to assess the efficiency of investigations when using Insight, in comparison to investigations conducted using Autopsy.

During this study you will be asked to complete a number of tasks using Insight or Autopsy. Please complete these tasks to the best of your ability. You will also be timed during this exercise, however, there is no specific time limit.

You will find more information in the Instruction Sheet. Please read this carefully before you begin.

Background

These tasks will use the John Doe test case which you are already familiar with. A Windows 7 virtual machine will be provided to you. This VM should be used when using Insight or Autopsy.

Requirements

To take part in this study, you must be either a 4th Year or Masters student. If you are colourblind, please mark the box at the top of the task sheet. This is important as it may affect your ability to use the application. If you struggle to use the application because you are colourblind, please indicate this in your answers.

Confidentiality

No personally identifiable information will be collected during this study. If you are colourblind, this information will be linked to your random participant ID.

You are free to withdraw from this study at any time, no reason needs to be given for withdrawal.

Should you wish to withdraw, please follow the withdrawal instructions given on the Instruction Sheet.

Contact

You are free to take this sheet with you. Should you have any questions or concerns after you have left, please feel free to email me at g.hales@abertay.ac.uk. If you would like a copy of Insight, or access to the source code, please ask.

Insight Usability Study - Instruction Sheet

YOUR PARTICIPANT ID IS:

YOU ARE REQUIRED TO USE ONLY INSIGHT / AUTOPSY 3

PLEASE READ ALL OF THIS SHEET CAREFULLY BEFORE BEGINNING

I am colourblind: ☐ YES (Mark Box)

To Begin

Feel free to read through the attached task sheet first.

Firstly, start the provided VM. Once the VM is started, start the 'TIMER' application from the Desktop.

Enter your Participant ID (shown above), click 'Start Test' when you have read this whole sheet and you feel you are ready to begin.

Tasks

IMPORTANT: Please make sure you have started the TIMER application, entered your participant ID and clicked 'Start Experiment' before beginning the tasks. Failure to do this will mean your effort is wasted. Do not close the TIMER application at any point, and do not restart the VM. If either of these things happens, alert Gavin immediately.

You can access the files taken from John Doe's computer by using the shortcut on the desktop marked Dataset. Please ignore all MAC times when using this folder, they are not forensically sound. The MAC times in Insight / Autopsy are correct.

You may access the internet, but please **DO NOT** use the opposite application to the one shown above (i.e. Do not use Insight if Autopsy is indicated at the top of this sheet, and vice versa). Please make a note of any other software you use on the task sheet. Please **DO NOT** use the timeline feature of Autopsy 3.

Work through and attempt all of the tasks using the application indicated at the top of this sheet.

Once you have finished, and are happy with your answers, please click the 'End Experiment' button.

Now please fill out the Feedback section of the form.

Once you are finished, please leave your completed tasks on your desk.

Problems

As Insight is a new piece of software, you may experience some problems. If the application crashes, simply restart it and carry on. Make a brief note of any problems in the Feedback section of the form on the very last page (section 7.4).

If the VM or the Timer application crashes, alert Gavin immediately. This will need to be recorded as it will affect the outcome of the experiment.

Withdrawing from the study

Should you wish to withdraw from the study, simply click the 'End Experiment' button on the Timer application. You should then mark the top of the instruction sheet with the word "WITHDRAWN". Please detach the task sheet and take it with you if you have filled in any information already. You are then free to leave.

Tasks

1. Firefox

1.1. The suspect installed Firefox on their PC. Find the date and time of installation.

1.2 The suspect initially tried to download the Firefox installer from a non-Mozilla website. Which website was this?

1.3 On a scale of 1 to 5, how easy was it to complete this task with the assigned tool? (Circle one)

1
Very Difficult

2
Difficult

3
Neutral

4
Easy

5
Very Easy

2. Avian Images

2.1 The suspect is known to own a Canon EOS-1D camera. Find a photo of a bird taken with this model of camera. What is the filename and path?

2.2 When was this image last accessed?

2.3 On a scale of 1 to 5, how easy was it to complete this task with the assigned tool? (Circle one)

1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy

3. Anti-Forensics

3.1 The suspect has attempted to hide a ZIP file full of bird images. What is the filename and path?

3.2 List the filenames of two of the images inside this ZIP file.

3.3 On a scale of 1 to 5, how easy was it to complete this task with the assigned tool? (Circle one)

1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy

4. **EXIF Metadata**

4.1 A large number of the photos on the suspects device were taken using the same camera, what make and model?

4.2 On a scale of 1 to 5, how easy was it to complete this task with the assigned tool? (Circle one)

1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy

5. **University**

5.1 The suspect took their device to a university at some point. Which university?

5.2 How do you know this?

5.3 On what date did they take it to the university?

5.4 On a scale of 1 to 5, how easy was it to complete this task with the assigned tool? (Circle one)

1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy

6. Suspicious Event

6.1 On 2nd Feb 2005 at 4:32pm, the suspect did something on their device which could be considered to be suspicious. What did they do?

6.2 On a scale of 1 to 5, how easy was it to complete this task with the assigned tool? (Circle one)

1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy

**IF YOU ARE SURE YOU HAVE FINISHED UP TO THIS POINT AND ARE HAPPY
WITH YOUR ANSWERS, CLICK 'END EXPERIMENT' ON THE TIMER APP**

7. **Feedback**

7.1 In general, how did you find the application to use?

7.2 Were there any parts of the application you found frustrating or difficult to use?

7.3 Were there any parts of the application that you thought were very useful?

7.4 Any other comments? (Record app crashes here too)

Thank you for participating, please leave the stapled sheets on the desk, these will be collected later.

Appendix B – Accuracy Results

ID	Q1	Q2	Q3	Q4	Q5	Q6	
IA21317	Y	Y	Y	Y	Y	Y	
IA84380	Y	Y	Y	Y	Y	Y	
IA59403	Y	Y	N	Y	Y	Y	
IA42852	Y	Y	N	Y	Y	Y	
IA86385	Y	Y	N	Y	Y	Y	
IA26957	Y	Y	Y	Y	Y	Y	
IA42284	Y	Y	Y	Y	Y	Y	
IA25759	Y	Y	Y	Y	Y	Y	
IA67180	Y	Y	Y	Y	N	N	
IA42267	Y	Y	Y	Y	Y	Y	
IA41979	N	Y	N	Y	N	N	
IA92704	N	Y	N	Y	Y	N	
IA60022	Y	Y	N	Y	Y	Y	
IA00564	N	Y	N	Y	N	Y	
IA36832	Y	Y	N	Y	Y	Y	
AA74402	N	Y	Y	Y	N	Y	COLOURBLIND
AA55372							NO TIME
AA82873	N	Y	N	Y	N	Y	
AA29458	N	Y	Y	Y	N	N	
AA48115							NO TIME
AA32679	Y	Y	Y	Y	N	Y	
AA86297	Y	N	Y	Y	N	Y	
AA18880	Y	Y	Y	Y	N	N	
AA33069	Y	Y	Y	Y	N	Y	
AA35804	N	Y	N	Y	N	Y	
AA18059	Y	Y	Y	Y	Y	Y	
AA22207	N	Y	Y	Y	N	N	
AA30846							INCOMPLETE
AA10303	Y	Y	N	Y	N	Y	

	Q1	Q2	Q3	Q4	Q5	Q6	
INSIGHT	80.00%	100.00%	46.67%	100.00%	80.00%	80.00%	
AUTOPSY	54.55%	90.91%	72.73%	100.00%	9.09%	72.73%	
p-Value	0.1730799	0.4230769	0.1925642	1	0.0004598	0.6698151	
CORRECT INS	12	15	7	15	12	12	
CORRECT AUT	6	10	8	11	1	8	

Appendix C – Likert Feedback Results

ID	Q1	Q2	Q3	Q4	Q5	Q6	
IA21317	5	5	4	5	5	4	
IA84380	2	4	5	5	5	5	
IA59403	4	2	1	5	4	3	
IA42852	4	5	3	5	5	5	
IA86385	5	5	NA	5	5	5	
IA26957	3	4	NA	4	4	4	
IA42284	3	4	3	5	4	5	
IA25759	4	4	3	5	4	4	
IA67180	2	4	4	5	1	5	
IA42267	4	5	2	5	5	5	
IA41979	3	4	NA	5	NA	NA	
IA92704	2	5	3	5	4	5	
IA60022	4	4	NA	4	4	4	
IA00564	2	2	2	2	2	4	
IA36832	2	2	3	2	2	3	
AA74402	3	3	3	3	2	1	COLOURBLIND
AA55372	4	4	3	4	2	NA	
AA82873	2	5	2	5	3	3	
AA29458	3	4	3	5	1	1	
AA48115	4	2	2	5	1	1	
AA32679	4	2	2	5	1	2	
AA86297	4	2	4	4	2	2	
AA18880	3	4	4	5	1	2	
AA33069	4	5	3	5	1	3	
AA35804	1	3	2	5	2	1	
AA18059	2	5	3	5	3	2	
AA22207	1	4	3	5	NA	1	
AA30846	5	5	4	5	NA	NA	INCOMPLETE
AA10303	4	5	1	5	3	4	

RESPONSES (INS)	15	15	11	15	14	14	
RESPONSES (AUTO)	14	14	14	14	12	12	
INSIGHT MEAN	3.266666667	3.933333333	3	4.466666667	3.857142857	4.357142857	
AUTOPSY MEAN	3.142857143	3.785714286	2.785714286	4.714285714	1.833333333	1.916666667	
INSIGHT SD	1.099783528	1.099783528	1.095445115	1.060098827	1.292412345	0.7449463437	
AUTOPSY SD	1.231455852	1.188313053	0.8925823753	0.6112498455	0.8348471099	0.9962049199	
INSIGHT MEDIAN	3	4	3	5	4	4.5	
AUTOPSY MEDIAN	3.5	4	3	5	2	2	
	Q1	Q2	Q3	Q4	Q5	Q6	
MANN-WHITNEY U VALUE	101.5	99	68.5	112.5	19.5	6.5	
MANN-WHITNEY p VALUE	0.8915	0.7988	0.6432	0.6834	0.0007602	0.00005073	

Appendix D – Feedback Thematic Analysis

Theme	Example
Intuitive User Interface	“Very easy, great layout” “Easy, user friendly, clear and really useful” “Easy to use and visual which was nice”
Coloured Event Categories	“Colour coding was useful” “...colour coding worked well...” Most useful features section: “Colour code of events”
Event Filtering	“The filters and colour coding worked well...” “The filter made finding each required element very straightforward” Most useful features section: “Event filters / search function”

Table D.1 - Positive feedback themes for Insight

Theme	Example
Event Clutter	<p>“The timeline was also very ‘busy’ / full at some points, especially during web browsing.”</p> <p>“Display of events was a little clustered. Could have been laid out a little better.”</p>
Buggy Timeline Scrolling	<p>“Timeline navigation, although good on the whole, was slightly frustrating with return to start of range.”</p> <p>“The timeline kept going back to 2000 when I didn’t want it to.”</p>

Table D.2 - Negative feedback themes for Insight

Theme	Example
Intuitive User Interface	<p>“Relatively straightforward and intuitive.”</p> <p>“It was quite easy to use and the GUI was well-organised.”</p>
Filesystem Integration	<p>“It was easy to search through the files”</p> <p>Most useful features section: “...browsing the filesystem, ability to jump to filesystem for any given file...”</p>

Table D.3 - Positive feedback themes for Autopsy 3

Theme	Example
Unintuitive User Interface	<p>“Useful, but clunky to use.”</p> <p>“Terrible and confusing.”</p> <p>“Too simple interface and difficult to find what I needed.”</p> <p>“Awkward, unresponsive, non-intuitive.”</p>
Search Function	<p>“Search function could be more refined, but was useful.”</p> <p>“...to search dates it’s complicated.”</p>

Table D.4 - Negative feedback themes for Autopsy 3

Appendix E – Participant Completion Times

<u>ID</u>	<u>Time</u>	<u>Secs</u>
IA21317	00:26:20	1,580
IA84380	00:23:15	1,395
IA59403	00:38:39	2,319
IA42852	00:34:01	2,041
IA86385	00:54:34	3,274
IA26957	00:52:51	3,171
IA42284	00:50:46	3,046
IA25759	01:04:39	3,879
IA67180	01:14:27	4,467
IA42267	00:54:21	3,261
IA41979	01:23:23	5,003
IA92704	00:36:48	2,208
IA60022	00:36:03	2,163
IA00564	01:30:09	5,409
IA36832	02:02:38	7,358
AA74402	2:22:28	8,548
AA55372		
AA82873	00:53:32	3,212
AA29458	00:53:57	3,237
AA48115		
AA32679	01:26:13	5,173
AA86297	01:22:36	4,956
AA18880	01:10:34	4,234
AA33069	01:05:34	3,934
AA35804	00:47:29	2,849
AA18059	00:40:49	2,449
AA22207	00:38:38	2,318
AA30846		
AA10303	00:37:14	2,234