12-15-2012

Creation and Testing of a Semi-Automated Digital Triage Process Model

Gary DeWayne Cantrell

Creation and testing of a semi-automated

digital triage process model

By

Gary DeWayne Cantrell

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi

December 2012

Creation and testing of a semi-automated

digital triage process model

By

Gary DeWayne Cantrell

Approved:

David A. Dampier
Professor of Computer
Science and Engineering
(Major Professor)

Yoginder Dandass
Associate Professor of Computer
Science and Engineering
(Committee Member)

Nan Niu
Assistant Professor of Computer
Science and Engineering
(Committee Member)

Alfred Christopher Bogen
Adjunct Professor of Computer
Science and Engineering
(Committee Member)

Sarah A. Rajala
Dean of the James Worth Bagley College
of Engineering

Name: Gary DeWayne Cantrell

Date of Degree: December 15th, 2012

Institution: Mississippi State University

Major Field: Computer Science

Major Professor: Dr. David Dampier

Title of Study: Creation and testing of a semi-automated digital triage process model

Pages of Study: 111

Candidate for Degree of Doctor of Philosophy

Digital forensics examiners have a growing problem caused by their own success. The need for digital forensics is increasing and so are the devices that need examining. Not only are the number of devices growing, but so is the amount of information those devices can hold. One result of this problem is a growing backlog that could soon overwhelm digital forensics labs across the country.

One way to combat this growing problem is to use digital triage to find the most pertinent information first. Unfortunately, although several digital forensics models have been created, very few digital triage models have been developed. This results in most organizations, if they perform digital triage at all, performing digital triage in an untested ad hoc fashion that varies from office to office.

This dissertation will contribute to digital forensics science by creating and testing a digital triage model. This model will be semi-automated to allow for the use by untrained users; it will be as operating system independent as possible; and it will allow the user to customize it based on a specific crime class or classes. The use of this model will decrease the amount of time it takes a digital triage examiner to make a successful assessment concerning evidence.

Key words: dissertation, modeling, digital forensics, digital triage, evidence previewing

DEDICATION

I would like to dedicate this dissertation in memory of my father J.C. Cantrell who left us September 27th 2011. He wanted to see me finish this so much, and I know somewhere he is proud of me for doing so.

ACKNOWLEDGEMENTS

The author would like to express his sincerest gratitude to the following individuals and groups who provided support and guidance throughout this research work:

- My entire family without them the author would not be where he is today,
- The author's PhD advisor and academic mentor who has provided guidance over many years,
- The author's PhD committee and all of MSU's CSE instructors who contributed to the author's education,
- The Mississippi State Attorney General's Office (MSAGO) Cyber Crime Unit,
- Student workers Steven Whitmire, Evan Howlett, and Nicholas Mason,
- My editors Joan Runs Through and Erin O'Brian
- The administration at Dixie State College,
- All my many friends and acquaintances who showed patience as I went through this process.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Digital forensics is a young field with many opportunities for research and development. This dissertation shows the results of creating and testing a methodology for a sub-phase of digital forensics commonly known as digital triage.

Digital triage is a process that occurs prior to the actual examination of the digital evidence in computer forensics. It is used as a technique to gather intelligence quickly when the situation calls for it. Digital triage has different uses depending on where it is utilized. It can be used in the field to help guide the search and seizure process, in the office to help determine if a piece of media is worth submitting for analysis, or in the digital forensics laboratory to assist in case prioritization.

There have been very few efforts to date attempting to create a digital triage process model. This dissertation presents the creation and testing of such a model. The model will be designed, each phase will be tested, and then it will be determined if the use of the proposed digital triage processing model can increase the efficiency of the digital triage process.

The hypothesis of the proposed dissertation is that the use of this model during digital triage will decrease the time it takes to perform digital triage, as digital triage is defined in this proposal, by at least 50% without decreasing the accuracy rate of decisions made during digital triage.

This chapter will define computer forensics, define the digital triage process, and describe further the hypothesis tested in this dissertation.

1.1 Digital Forensics

Digital forensics involves the post-event processing of a piece of digital media for artifacts of interest. An event in this case means a crime against a computer, a crime where a computer was a tool, or a crime where the computer is incidental to the crime[39]. This analysis usually involves the examination of the media for artifacts that can be used in a court of law. Since these items are to be used in a court of law, they have to be gathered using proven, forensically sound methodologies. Present day, these proven methodologies are typically standard operating procedures created independently by each office. Several models have now been written to better explain digital forensics and to attempt to make the digital forensics process more universal. Many of these models are based on or enhancements to the framework proposed by the first Digital Forensics Research Workshop(DFRWS) held in 2001[46].

This first framework consisted of the following phases:

1. Identification: the identifying of all physical items that could contain digital evidence
2. Preservation: a stage and an umbrella activity aimed at physical and logical protection to prevent data alteration and challenges in court
3. Collection: the gathering of the evidence without altering it
4. Examination and Analysis: the processing of the digital evidence for digital artifacts of evidentiary value

5. Presentation: the presenting of the evidence in a format the investigator can understand
6. Decision: involves those conclusions made from the Presentation phase

Although other models have additional phases or label their phases differently, the basic process is still the same and may have been best simplified by Warren Kruse as the the three A's of computer forensics[39]. Evidence has to be *acquired* without being altered; it has to be *authenticated* to prove it was not altered; and it has to be *analyzed* without alteration. This and other forensic process models are presented in more detail in Chapter 2.

The forensic imaging and authentication in the preservation phase of the digital forensic process is the most costly phase. Kenneally et al. makes the statement that the imaging of every piece of media before examining it is time and money intensive and does not necessarily result in more reliable evidence than imaging just the digital artifacts of interest[35]. However, complete forensic duplication is standard practice among digital forensic laboratories. Forensic imaging involves the duplication of every piece of media seized prior to it being examined and the authentication of each duplicate through the use of a strong hashing algorithm. Hashing is a mathematical technique that translates a variable length input into a fixed length output. Any change in the input typically results in a drastic change in the output. This fixed length output serves as a digital finger print that can be used to verify the evidence has not been altered[8].

Images can be either made with the use of an examination computer or a handheld imager. The handheld imager is a specialized piece of hardware designed for the sole purpose of forensic imaging. Therefore, it is typically the faster option[47]. Consider the following handheld imager speeds as claimed by the manufacturers.

Table 1.1

Time Comparison for Handheld Forensic Duplicators Current as of June 2011

Manufacturer	Image	Speed	Time to Image 1TB
VOOM	VOOM HardCopy3	7.1 GB/MIN	Approximately 145 Min
Logicube	Talon Advanced	7.0 GB/MIN	Approximately 145 Min
Tableau	Tableau TD1	6.0 GB/MIN	Approximately 167 Min

According to Table 1.1, using a hardware imager even in the best circumstance, an examiner still cannot look at a 1 terabyte drive for at least 2.5 hours when following the currently accepted digital forensics methodology. After the imaging process is complete, there is still the long process of searching, organizing, and categorizing the data. This is assuming nothing goes wrong during the imaging process. At present there is no other accepted way to produce court admissible evidence. For a much more detailed comparison of handheld imagers and more real world speeds, see the work done by Jack Riley of Mississippi State University[47].

Since its inception, digital forensics has only grown more complex. Now digital examiner must not only process computer hard drives, but digital examiners must also examine cell phones, personal digital assistants, Mp3 players, and any device that can hold digital information. Also, the amount of data needed to be processed per device is constantly growing and is likely to continue to do so[2]. The traditional process, as just described, means that the evidence is not examined until it has all been duplicated, and as was briefly shown this duplication process is not a quick one. This lengthy procedure creates many problems for the digital forensic examiner including constantly working from a backlog and large amounts of down time while waiting for machines to duplicate media.

1.2 Digital Triage

One answer to the growing problem of increasing backlogs due to growing hard drive sizes is digital triage. Triage is a term used in the medical field meaning to separate patients into those that will survive with no aide, those that can be saved with assistance, and those that cannot be saved even with assistance. Triage is usually applied in dire circumstances when there is no feasible way of treating every patient. In digital forensics it has come to mean any process that occurs before the traditional collection and preservation phases of the digital forensics process. It involves the examination of the digital artifacts that are most likely to provide useful intelligence first and saving a complete examination for later.

The desired result of digital triage is quick intelligence, not necessarily court admissible evidence, although information obtained can be admitted in certain circumstances such as when only enough information is needed to seek a plea bargain or quick validation of evidence is required. Digital triage can be conducted on site during search and seizure to provide feedback to the search and seizure team, it can be conducted prior to submitting the digital evidence to a laboratory to determine if the time a complete examination will take is viable, or it can be performed in the laboratory as a tool for prioritizing case loads[16]. Consider the extreme case of a multi-terabyte drive that has already been wiped clean.

There is no reason to go through the lengthy duplication phase or send it off to a laboratory for examination if there is no information there to be found. Consider also a search and seizure of 20 multi-terabyte servers from a corporation. Would it not be better to determine which machines are the most likely to hold the evidence before undertaking the lengthy task of imaging each one?

Digital evidence is very fragile, and trust in digital evidence was, and in some sense still is, very low. Thus, the first digital forensics methodologies were created with extreme preservation in mind. Since then there have been many improvements in hardware and software tools that can be utilized to prevent changing source media during examination or previewing. Most importantly examiners now have hardware write blockers that can be applied in conjunction with examination software that guarantee no write commands go to the source media. However as always in the world of criminal justice and law, policies are slow to change. Processes will need to be created, tested, and accepted in court before examiners can veer too far from the current accepted methodology of duplicating every piece of media before examining it.

Digital triage can have many uses once widely accepted by the law enforcement community. K. Rogers et al. in their work point out that intelligence gathered during digital triage creates a feedback loop that can be used to guide investigators during search and seizure[33]. For example, what about a child pornography case where no immediate contraband imagery is located, but clues found during digital triage point to external storage such as optical media or external hard drives that have not yet been found? Digital triage is certainly useful in this situation to help guide the search and seizure process. The psychological effects of immediate intelligence during suspect interviews are important to consider as well. Yeschke points out that the suspect is most vulnerable those first few hours after being apprehended[50]. Perhaps the easiest way to locate the missing optical media and external hard drives in our fictitious child pornography case is to ask the suspect during their first interview where they are located.

Digital triage is commonly performed on both live and dead evidence. In a live evidence scenario the digital triage examiner is working on an active machine to extract data elements needed for the investigation. For example, in the case of a live server digital triage could be performed to prevent down time of the server or to perform selective extraction of evidence due to storage constraints[35]. Another important use of live extraction with a digital triage tool is when encountering a volume with full disk encryption. Extracting data while the machine is still active allows the examiner to extract un-encrypted versions of all files. Once the machine is turned off, all files become inaccessible unless the password can be obtained. In this type of triage situation evidence altercation is unavoidable, but as already discussed, with the proper documentation this should not be an issue to the courts.

Dead analysis is conducted on evidence that has already been powered off either because the computer housing it has been booted into a digital triage environment or it has been seized and powered down. In a situation where the computer has been seized and analysis software/hardware is available, the digital media can be removed from the suspect's machine and attached through a hardware write blocker to another machine for analysis. Another option with both a live or dead machine would be to boot it into a digital triage environment with a live CD/DVD or bootable USB media. Once booted into this safe environment, the evidence is essentially in a dead state, and although not as secure as it would be with a hardware write blocker, it is protected from alteration.

1.2.1 Digital Triage Modeling

There is a precedence and existing tools for doing pre-examinations but few written procedures. The FBI's Image Scan tool for law enforcement use only is one such example of a tool designed for digital triage. This tool allows for image and video file scanning prior to evidence duplication. With this utility the digital triage examiner can scan for these files and then extract a sampling to justify seizing the computer and requesting a warrant.

More recently commercial products have started to become available as well. IDEAL Corporation produces several hardware solutions for doing digital triage (http://www.ideal corp.com/). According to their online documentation their tool, consisting of software and hardware components, allows a user with little to no technical training to gather ”actionable intelligence within minutes.” ADF Solutions Incorporated has developed software for digital triage (http://adfsolutions.com/). They also claim to allow users with little or no training to quickly gather intelligence. Finally, AccessData, a long time player in the world of digital forensics, has introduced a new tool for digital triage (http://accessdata.com/products/co mputer-forensics/ad-triage). Their tool also promises to facilitate successful previewing by novice users.

These tools can be very useful. However, what is missing is the validation and verification of the digital triage process, and thus much like the digital forensics process itself, even with the use of these tools the intelligence gathering process is still mostly ad hoc or based on each examiner's experience and/or each office's standard operating procedures.

There have now been several published works attempting to codify a digital forensics process model. However although some mention the possibility or the need, none seem to include an actual digital triage process as part of their model. All the models typically concentrate on what happens to the digital evidence after it has been duplicated and authenticated. There have also been very few attempts to create a process model just for digital triage. One exception is the The Computer Forensics Field Triage Process Model[33]. This model along with the other digital forensics models will be further discussed in Chapter 2.

1.2.2 Hypothesis and Contribution

The hypothesis of this dissertation is that the use of the Semi-automated Crime Specific Digital Triage Process Model will decrease the time it takes to perform digital triage, as digital triage is defined here, by at least 50% without decreasing the accuracy rate of decisions made during digital triage.

The testing of this hypotheses will contribute to scientific knowledge by formalizing and providing verification for a process that is currently done today in an ad hoc fashion. Several attempts have been made to model the process of digital forensics [3, 4, 6, 11, 12, 20, 30, 41, 46], but although some mention the need for a preprocess like digital triage, none attempt to create one as part of their model, and few processes have been made that are designed purely as a digital triage model[33].

Even though there have been few attempts to create a digital triage process, the scientific community acknowledges that the ever increasing size of digital media is a problem for digital forensics and digital triage could be one possible solution[2, 16, 34]. The few attempts that have been made to create a digital triage process, or like routine, have not included an automated process and has offered very little in the way of guidance for different classes of crime[24, 27, 33]. They also seem to depend a great deal on the expertise of the examiner to determine where to look and what to look for. This leaves a need for more automated and easier to use models for first responders and personnel with limited forensic training.

If adapted, this model could increase the efficiency of digital triage and this dissertation would support its use in digital investigations by showing the usefulness of digital triage as well as the usefulness of the model itself.

Currently, as has been discussed, the standard operating procedure is to image every piece of media prior to examination. This is creating a slow down in digital investigations and contributing to the rising backlogs of digital examiners. For digital triage to be accepted it must be modeled, tested, and then accepted by the law enforcement community. This work could contribute to that effort.

CHAPTER 2

RELATED WORKS

This chapter reviews related works concerning digital forensics process modeling, digital triage modeling, digital triage, and analysis efforts useful or related to digital triage. There are very few works directly related to digital triage modeling, but there are published works indirectly related to it such as digital forensic process models, digital triage, user profiling, and case prioritization.

2.1 Existing Digital Forensics Process Models

For any process to be tested and proven it must first be modeled, and then that model can be agreed upon as a standard procedure. The digital forensic process is not there yet. Computer forensics is commonly carried out through the development of individual office standard operating procedures [7]. These works represent an effort to codify these procedures in the form of a model or framework to establish a more universal standard of operations. They were important to include in this review to demonstrate what work has already been done to model digital triage and to show that the model introduced herein is unique.

2.1.1 Digital Forensics Research Workshop (DFRWS) Models

The Digital Forensics Research Workshop is a non-profit volunteer organization started in 2001. Their primary goal is to partner with academics and practitioners in an effort to assist in guiding research and development of the digital forensics field. One of the primary efforts of their first meeting was to define a framework for digital forensic science. This framework was not necessarily intended as a useable model, but instead a guiding framework for future works[46].

The DFRWS model is very important as it was one of the first attempts at modeling digital forensics and was created with the intention of being a grounded framework. It has the basic structure described by Warren G. Kruse as the three A's of digital forensics Acquire, Authenticate, and Analyze[39]. Evidence is acquired without being altered; evidence is authenticated to prove it was not altered; and finally evidence is analyzed without disrupting this authentication. The framework for the DFRWS model, as already mentioned in Chapter 1, is the linear process shown here[46]:

1. Identification: the identifying of all physical items that could contain digital evidence
2. Preservation: a stage and an umbrella activity aimed at physical and logical protection to prevent data alteration and challenges in court
3. Collection: the gathering of the evidence without altering it
4. Examination and Analysis: the processing of the digital evidence for digital artifacts of evidentiary value
5. Presentation: the presenting of the evidence in a format the investigator can understand
6. Decision: involves those conclusions made from the Presentation phase

Even though this process was described as linear, it was mentioned that a feedback mechanism should be introduced into the model to be more true to real life. In this same line of thought, it was suggested that preservation should be classified as more of a guarded principle across forensic categories. Preservation is not an activity that occurs once. Preservation is a need for every stage of digital forensics[8].

Mark Reith et al. proposed the Abstract Process Model[41]. They describe their model as, an enhancement of the DFRWS model since it is inspired from it. They add the additional phases of Preparation, and Approach strategy after the Identification phase of the DFRWS model, and a phase called Returning the Evidence to the end. The remaining phases are named the same and are similar to the DFWRS model phases. Neither model has digital triage as part of the process.

Seamus O Cirardhuain attempted to expand both the Abstract Process Model and the Digital Forensics Research Workshop model with his model the Extended Model of Cybercrime Investigations[20]. His contribution was the expansion of the model to incorporate the flow of information. His expanded model included Awareness, Authorization, Planning, Notification, Search for and identify evidence, Collection of evidence, Transport of evidence, Storage of evidence, Examination of evidence, Hypothesis, Presentation of hypothesis, Proof/Defense of hypothesis, and Dissemination of information.

2.1.2 Physical Crime Scene Adaptations

In 2003 Brian Carrier and Eugene Spafford published their model in the International Journal of Digital Evidence[11]. Their main premise was to design a model using the proven theories and techniques used in physical investigations. As they state,

> ”It is therefore important that a process model for the digital investigation exists and that it easily interacts with the physical investigations that have long existed.”

Their model contains 17 phases organized into 5 groups listed here along with the phase’s primary goal.

- Readiness Phases: an on-going phase involving the preparedness of the facility to handle digital investigations
- Deployment Phases: provides a mechanism for the incident to be detected and confirmed
- Physical Crime Scene Investigation Phases: processing the physical evidence at a crime scene including the physical media that holds the digital evidence
- Digital Crime Scene Investigation Phases: processing the digital evidence
- Review Phase: looking for ways to improve the process by reviewing the investigation

The sub-phases that can concern digital triage are located in the Digital Crime Scene phase.

- Preservation Phase: securing the digital evidence physically and logically as well as performing any volatile memory extraction, live forensics, if called for
- Survey Phase: searching the digital evidence for the most logically useful information such as digital images in a child pornography case (this phase can occur on a live system in the field making it essentially digital triage)
- Documentation Phase: The documentation of all evidence found
- Search and Collection Phase: More thorough examination using the information obtained from the survey phase

- Reconstruction: using advanced analysis techniques and experience to reconstruct events on the computer
- The Presentation Phase: provide the results back to the physical investigation team

The overall goal of the Brian Carrier and Eugene Spafford model is to treat the digital crime scene as a separate physical crime scene that feeds back into the real world. The Survey Phase of this model along with the Search and Collection phase have connections to digital triage. The survey phase if done in the field would actually be digital triage as it is being conducted on the evidence prior to the evidence being forensically duplicated, and if this phase was conducted in the office or in a forensics laboratory on the forensic duplicate, then the only difference between this phase and digital triage is that it is being carried out on a duplicate instead of the original. In both cases the evidence is being previewed in such a fashion to guide an examiner through a more thorough examination later.

In 2004, Benansius Baryamureeba and Florence Tushabe sought to improve on Brian Carrier and Eugene Spafford's model[11] with their Enhanced Digital Investigation Process Model[3]. Their proposed groups were slightly different and allowed for an iterative process instead of a linear one. However, the digital crime scene investigation phases were the same as Brian Carrier and Eugene Spafford's model[11]. This creates the same situation as far as digital triage is concerned. It is suggested as a need, but not described.

In 2004 Brian Carrier and Eugene Spafford presented another similar model[12]. They replaced the Review Phase with one they called the Presentation Phase where the results of the investigation are presented instead of the investigation process being reviewed. Once again the important grouping is that of the digital crime scene phases. These sub-phases are Preservation and Documentation, Digital Evidence Searching and Documentation, and Digital Event Reconstruction and Documentation. The first step taken with the digital evidence is that of preserving it which involves making a forensic duplicate and authenticating it with a secure hash value. No room is left in this model for digital triage making digital triage an optional event that would occur prior to digital crime scene investigation. However, the authors of this work support the idea of digital triage with their statement,

> ”We have become used to the concept that an image of the disk exists, but this concept may need to be reconsidered as disk sizes get larger and it becomes infeasible to make copies of every disk. pp. 7”

2.1.3 Hierarchical Based Framework

Nicole Beebe and Jan Clark proposed a model they called A Hierarchical, Objectives-Based Framework[4]. This model used inspiration from many of the models already discussed[11, 20, 41, 46]. Their main contribution was the division of the phases into first tier, principles, and multiple sub-phases. Their first tier phases were:

- Preparation
- Incident Response
- Data Collection
- Data Analysis
- Presentation of Findings
- Incident Closure

Preparation is the same as the readiness and planning phases already discussed. Incident Response involves developing a game plan and validating the suspected incident did occur. They are not entirely clear, but it seems some digital evidence can be gathered at this point from the suspect's machine, but not a full examination or full collection. This is as close as they come to a digital triage stage. The Data Collection Phase is where the complete collection of all digital devices of evidentiary value would be collected, authenticated, and transported. Data Analysis Phase is where the actual examination is performed. Presentation of Findings provides for the interpretation of the evidence into forms that can be used in court presentations, reports, etc. This is the same as the previously described presentation phases. Incident Closure is the same as the review phase from Brian Carrier and Eugene Spafford's model already described[11].

Their main complaint with existing models is that most models to date are single-tier, higher order process models. They use the analogy of flying an airplane being a function of take-off, fly, and land. They point out this is not sufficient enough of a model and they propose their model as a more comprehensive digital investigation framework that allows for expansion with their hierarchical structure even though they do not actually elaborate on what most of those second tier sub-phases would contain. One improvement this model did make was to remove those things that should be part of every phase and call them principles. Good documentation is an example of a principle. Quality documentation should be created and maintained through out all phases.

A digital triage process was not described, but would fit into the incident response phase as part of the initial evidence gathering process. It is interesting that they never use the term digital triage, but they are describing the need for such a separate process very well.

> ”Some data and information will be initially collected during the Incident Response Phase, which is necessary to validate the incident and determine the impact and nature of the incident. The formal Data Collection Phase, however, occurs subsequent to the Incident Response Phase once a decision has been made that a digital investigation will ensue, regardless of its scope or anticipated legal or administrative actions.”

The first model mentioned was that of the DFRWS framework. In describing this model it was stated it follows the Warren G. Kruse pattern of the three A’s of computer forensics Acquire, Authenticate, and Analyze. All of the models discussed so far are laid out in this same linear fashion. None explicitly describe a digital triage process, but a few did describe the need or include a sub-phase where it would clearly fit in. In seems clear that all models agree for digital evidence to be admissible it must be acquired and authenticated before being analyzed. This leads to the standard operating procedure of not looking at any evidence until has been imaged first.

This does not mean analysis can not be done for quick intelligence. The key is the testing of the tools that are used to do the intelligence gathering to prove they do not alter evidence and the documentation showing that the pre-analysis, digital triage, was done.

2.1.4 Outside the DFRWS Framework

Other models were examined that did not as easily relate to the structure proposed by the DFRWS. In these models digital triage would fit in, but their processes were less of a linear procedure and more of an alternate approach or a concentration on one particular activity. A digital triage process was not proposed, but not excluded either. These models are mostly independent of each other.

Christopher Bogen and David Dampier, using inspiration from software engineering and the Unified Modeling Language views, suggest the separation of digital forensics modeling into three different views: Investigative Process View, Case Domain View, and Evidence View[6]. Their model concentrates on the actual examination phase. Therefore, digital triage would be a task prior to their model. The inclusion of a digital triage stage would not, however, interrupt their model.

Another model that introduced a process concentrating on the examination phase was that of Gong Ruibin et al.[48]. In essence they were trying to describe what was already going on instead of proposing a new model of action. They described the examination phase as being carried out through the following cyclic steps.

1. Seed Information Search
2. Data Filtering
3. Decision Making
4. Information Extraction
5. Query Refinement

Essentially an investigator develops a list of keywords to start the investigation. He or she then filters the data based on those keywords, examines the results, and refines that keyword list.

Their purpose in describing this model was so they could introduce an automated case relevance information extraction process. Case relevance being a concept they proposed. They define this concept as, ”the property of any piece of information, which is used to measure its ability to answer the investigative ”who, what, where, when, why and how” questions in a criminal investigation.”

They further provide a scale for case relevance ranging from Absolutely Irrelevant to Probably Case-Relevant with various levels in between. Their described automated case relevance information extraction routine is as follows:

1. Survey Sub-Phase: the creation of a case profile by a user to include a set of seed words to begin the search
2. Extraction Sub-Phase: a fully automatic process that starts with a few keywords and retrieves more and more case relevant data in an iterative fashion by increasing the search results with synonyms of the current search words
3. Examination Sub-Phase: the examination of the results by an investigator to interpret the evidence and determine if further extraction is required.

They go on to further explain how this iterative process would work. The relevance to this dissertation is that it describes another process that does not include a digital triage stage. Processing of the evidence occurs first, and there is no pre-examination phase. Secondly, an automated routine like this would be very useful in a digital triage tool if it could be done in a timely manner. However, their routine requires large amounts of comprehensive word searches which limits its usefulness to digital triage as those searches would take a considerable amount of time with current technology.

The Ricci framework FORZA proposed in 2006 is organized more into roles than actual events or actions[30]. These roles include:

- Case leader
- System/business owner
- Legal advisor
- Security/system architect/auditor
- Digital forensics specialist
- Digital forensics investigator/system administrator/operator
- Digital forensic analyst
- Legal prosecutor

For each role there is a set of six key questions what, why, how, who, where, and when that helps guide the tasks of that role.

This model deals less with the process of specific events and more with the organization of tasks. There is no inclusion of a digital triage process. However, in this situation the digital forensic specialist would decide the strategy which may or may not include a digital triage process. It would be up to the digital forensics investigator to carry out the process if needed, and either the investigator or the examiner to interpret the results. The information provided by the triage process would help answer the six key questions, and could be used to guide the rest of the steps. The model proposed for this work could be used in that optional digital triage phase.

2.2 Digital Triage

The following works are those that provide support for, model, or discuss the need for digital triage. They served to provide a deeper understanding of the need for triage, to help show support for its need and use, and to examine how it is currently being used.

2.2.1 A Call for Digital Triage

One of the efforts of the Colloquium for Information Systems Security Education (CISSE) in June of 2008 was to develop an outline for future research in digital forensics. After deciding that law and education deserved separate consideration, they created categories for digital forensics research.These categories were:

- Evidence Modeling
- Network Forensics
- Data Volume
- Live Acquisition
- Media Types
- Control Systems

Two of their categories were of particular interest to this work. The call for evidence modeling closely matches the crime potential templates described in the digital triage model used in this dissertation. They call for the modeling of cases for specific crime classes. Under the Data Volume category they also mention the need for research into tools that image and look for evidence at the same time. This is not a specific call for digital triage, but the process described herein could be applied during imaging for case prioritization if the technology was developed to do so.

In 2009 E. Casey et al. called for a departure from the traditional approach to seizing digital evidence in their paper Investigation Delayed is Justice Denied: Proposals for Expediting Forensic examinations of Digital Evidence[16]. Referring to the traditional digital forensics process where each piece of media is seized and duplicated before being examined, they state, This approach is rapidly becoming untenable as more cases involve increasing quantities of digital evidence. They propose that digital examinations be divided into three categories:

1. Survey/triage forensic inspection
2. Preliminary forensic examinations
3. In-depth forensic examinations

Category 3 is well documented and can be considered the traditional forensic process. The models discussed in the Digital Forensic Modeling section are good examples of category 3. This process involves the collection and duplication of every piece of media that could possibly contain digital evidence. This a very lengthy and time consuming process. They are proposing to only use this lengthy traditional process when it has already been determined that a piece of media will contain evidence. This determination being made by conducting preliminary examinations as described in categories 1 and 2.

As they describe it, the category 1 examinations determine if something of evidentiary value is present on the media being examined, and category 2 translates that evidence into something the investigator can use. They do not actually describe a process model for what should happen in category 1 and 2 exams, but describe how it can be useful and suggest a few commercial tools. According to Casey et al. a lot of offices are already doing digital triage in some form, but there are no published works or standard operating procedures for digital triage leading to every office having no choice but to develop their on method [16].

Personal computers are not the only device that can be triaged. For example, in 2010 Chevonne Dancer, of Mississippi State University, describes the need for a digital forensics process model for smart phones, and makes a call for the development of a process model concentrating on those traits that are the same for all smart phones [21].

More specific to digital triage, that same year Richard Mislan et al. took the categories mentioned by Casey et al. just described [16] and attempted to expand the work by describing the need for the category 1 and 2 examinations on mobile devices instead of computers[42]. They point out that with the growth in cell phone and mobile device use, digital forensic laboratories are becoming swamped with requests for mobile device evaluation. This overload of requests could be greatly reduced through the triage of mobile devices to determine which mobile devices should be sent in for analysis as evidence, which ones contain no evidence, and which ones could just be used for intelligence purposes. They also discuss the usefulness of performing mobile triage on the scene to guide the investigator, an idea that has been mentioned by other works in support of digital triage. Unfortunately triage technology for mobile devices is lacking especially when searching for tools that perform push button forensics.

Richard Mislan et al. offers a general set of guidelines for triage inspections instead of a process model, and proposes a minimum set of requirements that also serve as good guide lines for creating a digital triage tool for any media. Those rules are summarized and paraphrased here with an emphasis on the applicability to this work.

- user-friendly to accommodate the less advanced user
- inform the user of what can and cannot be obtained from the device
- make no changes or inform the user if an action might cause a change
- acquire data accurately
- include security to prevent unauthorized use
- provide meaningful errors
- maintain a complete log

Their general guidelines or, ”standard operating procedures for handling sources of digital evidence” are as follows.

1. initiate chain of custody
2. isolate device from network (if feasible and applicable)
3. disable security features (if feasible and applicable)
4. extract limited data
5. review extracted data
6. preview removable storage media

Their work was a call for more work in the triage area, but that call was aimed at mobile devices. During that call they propose a process for extracting data from mobile devices and a set of goals for a tool that would perform that extraction. That process and those goals serves as support of this work, but not a replacement.

Erin et al. make a call for an alternate approach to data collection in their paper, *Risk Sensitive Digital Evidence Collection*[35]. They make the case that the current methodology of imaging all of the media before viewing it is not only time and resource draining, but that changing the procedure does not necessarily result in less reliable evidence. There just needs to be a change in the accepted methodology.

Their solution involves the selective extraction of artifacts during the imaging process of digital forensics instead of the creation of a complete duplicate. Prior to this extraction those artifacts have to be identified. They do not go into detail on how those artifacts are chosen. In fact in their follow up paper at the DFRWS 2005 they make a call for research into the creation of, as they describe it, ”formal protocol or artifact identification (platform agnostic).” The proposed model in this work could be used as an identifier for that selection.

Their work does not suggest digital triage as a solution per se, but they are describing the need for a change in current methodologies which could include digital triage. If selective extraction were to become an accepted method of digital evidence collection, this work would prove very useful in the determination of the artifacts to be extracted. Their process is designed to be part of the collection phase, post digital triage. Although this work could be beneficial to their effort, the two efforts are mutually exclusive.

2.2.2 Digital Triage Process Model

None of the models discussed so far described the phases of a digital triage process model. However, one separate digital forensics triage model was identified as the Computer Forensics Field Triage Process Model[33]. Their model consists of the following phases:

1. Planning
2. Triage
3. Usage Profiles (Home directory, file properties, and registry)
4. Chronology/Timeline
5. Internet (Browser Artifacts, Email, IM)
6. Case Specific

Phase 1 is simply proper planning and preparation. Although not included in the original DFRWS model[46], this type of phase was proposed in many of the models that followed it and that have already been discussed[4, 3, 11, 12, 20, 41]. The triage phase is similar to the category 1 exam described in[16]. Evidence is previewed and prioritized based on what is likely to have the largest quantity or most valuable evidence. The remaining phases of this model more closely resemble category 2 examinations as described by[16]. The Usage/User Profiles phase involves the examination of items that can potentially be tied to an individual by concentrating on items tied to specific user accounts. The Chronology/Timeline phase is an attempt to sort and quantify the evidence based on modified, accessed, and created time stamps. The Internet phase involves the examination of items that pertain to Internet activity such as email and browser artifacts, and the Case Specific phase involves the examination of items that would typically be associated with the case in question. For example, one could concentrate on finding images in a child pornography case.

In 2002 Special Agent Jesse Kornblum described the need for digital triage tools and described several primary goals that would be appropriate for such a tool. One of those goals was that a first responder tool should do all the work without user intervention as first responders will often be untrained[38]. The Computer Forensics Field Process Model is described as a field model indicating its primary users will be first responders. However, the types of searches described in this model would require a lot of expertise. The authors even mention this at one point when they say, ”A thorough knowledge of user profiles and artifacts relating to usage, are essential to accomplishing this goal.”

This model also seems to depend on the presence of an operating system as indicated by the search for user profiles, and Internet artifacts. These items will typically only be on the main operating system drive. If the media being examined is external storage, then only phases 1, 2, and 6 are actually useful. Also, different operating systems are going to have much different user profiles and store Internet artifacts differently. Therefore, it is dependent on the first responder having expertise in user artifacts, but that expertise has to also be in sync with the operating system in question. These weaknesses were kept in mind when designing the model tested in this dissertation.

2.2.3 Dangers to Digital Triage

Although not specifically about digital triage, G. Bell, and R. Boddington performed a series of experiments that could have a direct effect on digital triage[5]. It appears the next wave in hard drive technology is likely to be solid state drives according to G. Bell and R. Boddington[5]. Solid state drives (SSDs) have no moving parts. Therefore, they are faster, less error prone, and more energy efficient. They are faster, but unlike magnetic drives there is a time penalty for writing over areas that have already been used as they have to be erased first. Some of these SSD drives have started using idle time to erase unused sectors. These erasure routines reside in the firmware of the drives. Being in firmware, there is no way to interrupt the process with software.

This only affects deleted data. Therefore, it has no appreciable impact on the process proposed in this dissertation. A digital triage process is typically going to concentrate on present files as deleted files and file fragments take time to locate and extract. Therefore, the effect on information retrieval is very low, and the digital triage process takes place prior to the authentication phase leading to no effect on authentication.

This erasure process does occur just by the drive being powered on. Therefore, the digital triage examiner is altering the evidence just by powering a hard drive on, and could be destroying evidence located in unallocated space. The actual likelihood of this was unclear from their experiments, as the only drives tested were those that had been formatted. This means they did not test it on files that were recently deleted or drives that are partially full.

Their tests also seem to indicate that the erasure process occurs quickly and is complete within hours. This means that if the drive is being regularly used then there is likely no evidence in unallocated space anyway. Finally, a case will typically not depend on files that have been carved as they have no system metadata(ownership information, time stamps, security information, etc), and although recently deleted files are admissible the best evidence items are those files that are present and being used. This study was further supported later by Christopher King and Timothy Vidas in their article Empirical Analysis of State Disk Data Retention When Used with Contemporary Operating Systems[37].

These works have no appreciable effect on this proposed process, but a digital triage examiner should be aware of the issue. They can easily overcome this problem by being aware that the problem exists, acknowledging that they are going to change the evidence, producing proper documentation showing it was necessary, and being able to explain the effect in court. Their experiments were based on if it happened. They were not based on how it happened. Once the mechanism for doing this erasure is understood, then hopefully ways for overcoming it can be developed and it will be avoided as an issue altogether.

2.3 User Profiling and Case Prioritization

A lot of work has been done on user profiling in the context of data mining for commercial use, and a lot of work has been done on user profiling as it pertains to network attacks and incident response. The concept of learning about users, as will need to be done in the Computer Profiling stage of the model discussed in Chapter 3, is more limited. The following works were insightful as to what has been done and what can feasibly be done in the proposed model in relation to computer/user profiling.

2.3.1 User Text Classification

One of the first areas explored in this part of the review was the idea of what information can be gathered just by analyzing the text on a suspect's machine. These papers explore the idea of automated authorship analysis. Authorship analysis is not a new field as applied to the written or typed word, but attempts are now being made to bring it into the digital forensics world. Some of the areas that have been explored include biometric analysis (determining a user by their keystroke patterns) qualitative analysis (based on the examiner's experience) and stylometry (focusing on countable language features and idiosyncrasies)[18].

Patricia Carter in her work attempted to use login session data alone to identify a user[13]. Patricia combined the logon session data, including which windows were opened and which programs accessed, into a document. She then applied already successful authorship analysis algorithms to the problem of identifying which user was logged on from a group of users.

Carole Chaski optimistically wrote about the use of authorship attribution in digital investigation to determine the actual individual who was using the computer at the time the evidence was created, a common concern in digital forensics[18]. She discusses the problems with using authorship analysis in the digital world, and demonstrates it is possible for it to be used by describing several experiments and through a discussion of examples cases where it has been successfully used.

More recently there have been publications on the application of authorship analysis for specific types of digital evidence such as instant messenger logs[45], and the microblogging website Twitter[40]. These works concentrate on finding solutions for the problem of performing authorship identification on limited sets of data.

Another effort worth mentioning is being undertaken by James Stinson of Mississippi State University. His work, which is yet to be published, involves the authorship analysis of computer programmer code in an effort to identify the author. This could be used, for example, in academic dishonesty detection of computer science classes or in tracing down the creator of malware. For this model, it could be applied in hacker scenarios by attempting the authorship identification of code found on the machine in question. However, its use in digital triage is limited by time and the availability of samples similar to the other analysis techniques discussed.

Authorship attribution, although seemingly useful to digital forensics, does not appear to be useful to this effort, and will not be explored further. The reviewed works indicate the primary goal of authorship attribution is to identify a set of writing as coming from a particular person typically using a pool of works from a set of likely individuals, and can be a lengthy process involving training and comparison. There were variations from this researched in the form of trying to identify some specific characteristic about the individual such as the gender of the author[49]. This type of focused information might be useful in digital triage if it can be performed in a timely manner, but it was still determined that, in general, authorship attribution is not appropriate to this work.

2.3.2 Automated User Profiling

Simson Garfinkel ran a set of experiments using what he termed pseudo-unique identifiers like social security numbers, credit card numbers, and email addresses to perform cross-drive analysis[24]. Cross-drive analysis in this case being the examination of drives to try and determine what they might have in common and automatically identify drives of interest. For example, identifying a set of hard drives which came from a set of users who are part of the same social network or work for the same corporation. During his experiments he also examined single drives to try and determine if they had information that violated user privacy and that should have been wiped before the drive was disposed.

In this work he also tried to determine who the primary user was based on their email address. This was accomplished through the creation of a histogram of email addresses found on each machine to determine the primary user based on the theory that the user's email address will appear more often than anyone else's. His work was designed to be used for case prioritization, but not necessarily during digital triage. However, the idea of pseudo-unique object extraction, and primary user determination through the use of email histograms could be very useful in computer profiling during digital triage.

A work published in 2009 entitled, *Fast User Classifying to Establish Forensic Analysis Priorities* comes closer to the goals of this dissertation[27]. In this work the authors describe what they call the Five-Minute-Forensics technique. They first extract a limited amount of information from a device, and then use it to rank the disks in question. Their goal is to decrease the time to choose between multiple hard drives during an examination and to avoid wasting time on data that is not useful. They first choose a small subset of files to extract including user specific registry files and Web cache. Then, they use this data along with the types of files found on the system to attempt to classify the user as: occasional user, chat-internet user, office worker user, experience user, and hacker user. The classification is determined by first training their system with hard drives of manually pre-classified user hard drives. Where as this work served as inspiration, it does not supplant this dissertation effort as their model requires training. Time was not a consideration as their goal did not involve digital triage.

2.4 Summary of Reviewed Works

This literary review looked at existing process models and discovered that no current model includes a digital forensics triage process, but that the inclusion of one will not interrupt current accepted practice. This review also searched for stand alone digital triage process models and only discovered one the Computer Forensics Field Triage Process Model[33]. The concept of digital triage was also explored in current literature, and it was shown that researchers are seeing the need for a digital triage process and call for research in this area.

Support for the automated computer profiling phase of the process model was searched for, and it was discovered there has been some work in user/computer profiling. However, most existing techniques require either too much user intervention or were not intended for use in digital triage and are thus too complicated or too time consuming for its use. However, these works did provide good inspiration and were useful during model development and testing.

CHAPTER 3

DIGITAL TRIAGE PROCESS MODEL

The primary contribution of this dissertation is the testing of a digital triage process model. This chapter will describe that model in detail. First, an overview of the model will be given. Then, a detailed description of each phase will be provided. Finally, a description of how it could be used will be discussed. Chapter 4 describes the creation of the tool that implements the automated portions of the model, and Chapter 5 will provide detail on how this model was tested through the tool introduced in Chapter 4.

3.1 The Model

The model introduced in this chapter is a semi-linear framework. The model attempts to allow for use by novices, incorporate ongoing concerns, and allow for specialization for different crime classes. The most significant contribution of the model is its automated phases shown in grey on Figure 3.1. Planning and Readiness is an on-going phase that occurs pre-event, and Preservation is an umbrella activity that is needed throughout all phases. The remaining phases are presented in a linear fashion. The dotted line on the left of Figure 3.1 represents the information flow from the Computer Profile and Crime Potential phase into the Presentation phase where it is transformed into usable information for the digital triage examiner. Each phase will be described in detail.

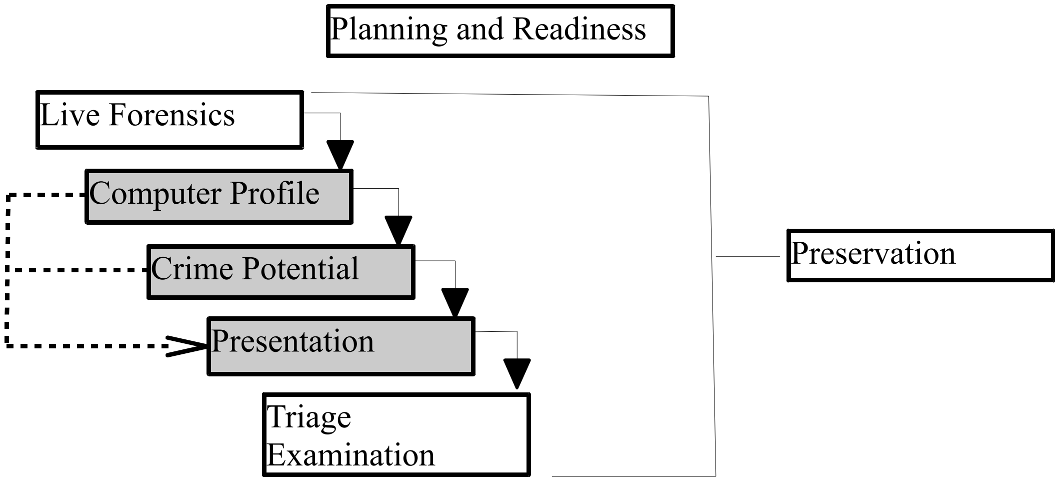


Figure 3.1

Semi-automated Crime Specific Digital Triage Process Model

3.1.1 Planning and Readiness

Planning and Readiness is an ongoing phase involving the preparation and education of staff, and the continual upgrading of equipment. Including a phase like this in a digital triage process is particularly important as it involves the continual testing of the triage tools and the effort to stay current on all hardware that the triage examiner may encounter.

The need to stay current with technology will always be an important part of any digital forensics process. New technology can often leave booby traps for the uninformed. For example, in Chapter 2 a work was mentioned by G. Bell et al. In this work it was shown that some solid state drives will start to erase unallocated space without any user interaction disrupting the authentication process by being powered on[5].

This work has minimal effect on the model described here, but issues like this can damage a case if a digital triage examiner is not aware of them. The digital triage examiner could wipe space that a complete examination could extract data from, regardless of the write blocking technology applied, just by applying power to the suspect's drive. If the triage exam is conducted after the evidence is processed, the triage examiner could disrupt authentication. The digital triage examiner can record their actions properly and account for the loss of authentication later if aware of the danger. If completely unaware of this danger, it is very likely their authentication will be corrupted. Planning and Readiness involves preparing for these situations.

3.1.2 Live Forensics

Live forensics involves the acquisition of volatile memory, and it is included as a phase of this model as an optional step dependent on need and expertise. Digital evidence should be gathered in the order of volatility to prevent the loss of any data of evidentiary value[23]. Thus, this step has to come before anything else in any digital forensic process to preserve volatile memory, but it is presently skipped in most investigations and the volatile memory typically ignored. It is most commonly used during incident response to attempt a determination of how a machine was compromised. It is also commonly used on live evidence when full disk encryption is suspected to acquire evidence before the machine is powered down or for password cracking later. Therefore, live forensics is likely to become more important and more common as users become familiar with and operating systems come standard with full disk encryption[17, 28].

The reasoning for including it as a phase in this model is that once the digital triage process begins volatile memory will be lost. A digital triage tool is either going to require a system reboot into a safe environment using boot media which will destroy all volatile memory or the running of a program from an external drive which will cause partial damage. In either case volatile memory will have been altered. The input from this phase will not be applied to the automated process.

3.1.3 Computer Profile

Computer profile generation will be the first phase in the automated process for the digital triage process model. In the Field Triage Process model the authors attempt to learn about the users of a system by targeting the user profiles on the computer[33]. The Five Minute Forensics Technique performed a similar analysis, but used the information to categorize the users into occasional user, chat-internet user, office worker user, experienced user, and hacker user[27]. The Semi-automated Crime Specific Digital Triage Process Model attempts to refine these process models by stream lining the process, implementing it in an automated fashion, and including customization for different crime classes in the form of a red flag report.

This phase involves the gathering of three types of information. First, general file system data is gathered. This allows the examiner to quickly see what type of and how much data is present. Basic file system information can also be used to get a feel for the complexity of the drive, and possibly the complexity of the user. Next file type statistics are gathered for the complete volume and for the individual user directories when possible. This allows the examiner to theorize about the primary usage of the drive. Web history is then gathered in an attempt to reveal a user's interest, communications, and recent activity.

Finally in the case of Windows machines, the registry is searched for. The registry is a gold mine of quick useful information. More detail concerning this information gathering process is provided in chapter 4.

In the case of external storage or computers that are not connected to the Internet, the only results will be file system information and file statistics. Further searches could be incorporated involving the search of document metadata or other items of interest. However, at present, profiling of external media will provide less results for this model. This phase is summarized in Table 3.1. Summaries will also be provided for the Crime Potential, Presentation, and Digital Triage phase as they are the primary focus of the model.

Table 3.1

Computer Profile Phase Summary

Name:	Computer Profile
Description:	First automated phase, builds a general profile of the computer
Input:	Physical disks desired for analysis
Precondition:	Authorization to search the system (warrant, consent to search, corporate authorization letter etc.), system booted into a safe Linux environment
Postcondition:	Computer profile data stream created
Product:	Data stream of information collected

3.1.4 Crime Potential

The Computer Profiling phase just described will be the same for every piece of media examined. In contrast, the Crime Potential phase contains those components that are dependent on a specific crime class. It will attempt to guide the triage examiner by raising red flags for a specific crime class. Although listed as a separate phase, it runs concurrently with the computer profiling stage to save time.

Different crimes will call for different template contents. Child pornographers are likely to hoard and collect images leading to a large ratio of image files to total files on their machine[36]. Therefore, a large percentage of image files might be a red flag when performing triage on a piece of media. A hacker is likely to have a large number of scripts and executables. Both crime classes have certain commonly used key words associated with each that can be searched. Crime templates will need to be developed for each crime class. This template would help identify those items of particular interest in a given class of crime such as certain file types or particular key words. This type of crime class modeling has already been identified as an important area of research at the 42nd Hawaii International Conference on System Sciences[43].

This template could become more advanced over time to include additional processes and filtering per crime class or more advanced filtering could be incorporated depending on the time critical nature of the examination. For example, it could be useful to be able to apply flesh tone filtering to images during child pornography cases such as done by the first responder tool File Hound[19, 26]. For the purposes of this dissertation, a very simple template will be used to demonstrate the usefulness of the model involving filtering with keywords.

Table 3.2

Proposed Standard Crime Template

Keywords	words of particular interest in the crime class
File Type Alerts	file types that would normally be found on a computer for a specific crime class
Known File Alerts	known files to be of interest in a specific crime class

This crime template allows for the search for both known files, by name, and keywords. Some types of criminal subcultures such as child pornographers, hackers, or fraudsters have their own unique slang that can be searched for. There are also image files that are commonly collected by child pornographers that are part of unique sets, and certain hacker tools that might commonly be found on a hacker's computer. These files can be searched for by name or if time is available by the more robust method of mathematical hash value. Unless altered, these files will always produce the same hash value providing a way to red flag them no matter what they are named. These commonly used words and known file hash values are already being used by some digital forensics tools during analysis and examination. For example, the National Institute of Standards and Technologies(NIST) maintains the National Software Reference Library which includes hash values for standard programs and those that are malicious in nature[1].

File type alerts are another category covered by the crime template shown in Table 3.1. A preponderance of certain file types could also be a strong indicator of certain activity. A child pornographer is likely to have a lot of image files, a hacker a lot of scripts and code, and a fraudster a lot of numerical based files and financial records. These statistics can be further divided into categories of files such as audio, video, documents, etc to help with intelligence gathering[27].

When choosing or designing a template a triage examiner must take into consideration the location the triage will occur. Will it take place in the field, in the office, or in the examination laboratory? The digital triage examiner must consider the time critical nature of the case. The more extensive the template the longer the automated process is likely to take. The digital triage examiner should also consider any legal ramifications. A warrant allowing the digital triage examiner to search for evidence of a murder does not allow the search for child pornography as well. See Table 3.3 for some example templates. See Table 3.4 for a summary of the Crime Potential phase.

Table 3.3

Example Crime Templates

Example Template	Child Pornography
Keywords	commonly used words by child pornographer collectors, known victims, known child porn image names
File Type Alerts	Image files
Known File Alerts	Known child porn images
Example Template	Murder Investigation
Keywords	words associated with the case
File Type Alerts	none
Known File Alerts	none
Example Template	Suspected Hacker
Keywords	known hacker words and phrases
File Type Alerts	scripts, executables
Known file Alerts	known root kits, known hacker tools, non-standard user software

Table 3.4

Crime Potential Phase Summary

Name:	Crime Potential
Description:	Second automated phase, uses key words to filter the computer profile information created in the previous phase
Input:	Data stream from the Computer Profile phase
Precondition:	Computer Profile phase completion
Postcondition:	All data filtered and organized
Product:	Additional data stream of filtered information

3.1.5 Presentation

In this phase the information from the User Analysis and the Crime Potential phase will be translated into a report that can quickly guide the digital triage examiner to items of interest or help quickly determine how the evidence should be prioritized. The results will be interpreted and applied according to need. If detailed enough, the examiner may be able to skip the next phase entirely saving time and effort. The utility created to test this model creates a simple HTML report divided into a main report that is unfiltered and a red flag report that only contains those items filtered by the Crime Potential phase. HTML reports are common to digital forensics examiners already and provide a dynamic environment useful for viewing evidence. More detail about the contents of the report can be found in Chapter 4, and a report sample can be found in Appendix A. See Table 3.5 for a summary of the Presentation phase.

Table 3.5

Presentation Phase Summary

Name:	Presentation
Description:	The third and final automated phase, organizes all reporting information
Input:	Data Streams from Computer Profile and Crime Potential phases
Precondition:	Crime Potential phase completion
Postcondition:	Report generated and presented to the user
Product:	HTML report separated into a main report for the Computer Profile phase data stream and a filtered report for the Crime potential phase data stream(see Appendix A for an example)

3.1.6 Triage Examination

The triage examination phase is the viewing of the evidence in a forensically safe manner using the guidance provided by the Presentation phase. This phase is optional dependent on need. If the Presentation phase produced enough information, then there will be no need for further examination.

This phase is different from a traditional examination. Typically, after the evidence is duplicated in a traditional digital forensic examination, the forensic software will assist in indexing, sorting, and categorizing all files and file fragments. One analogy is that of a filing cabinet being emptied and all the files sorted and categorized into separate stacks so the examiner can quickly locate information. The triage examination occurs prior to this duplication and sorting. Essentially in this phase the triage examiner is viewing the file system as it would be presented to the user in a file system explorer application. Any sorting or searching has to be done in real time by the triage examiner. In triage examination the examiner is looking through the filing cabinet to see if anything is readily apparent guided by the information provided in the Presentation phase.

Triage examination is not a new concept, but the addition of an automated process using predefined templates specific to each crime class is a new concept. This triage examination phase would be what typically occurs when an examiner uses any tool on the suspect media for information gathering purposes prior to evidence duplication. The automation and the guidance provided by this model will, however, make this phase more successful or unnecessary if the desired information is already in the produced report. See Table 3.6 for a summary of this phase.

Table 3.6

Digital Triage Examination Phase Summary

Name:	Digital Triage Examination
Description:	This is a manual phase and not necessary if all desired information is present in the report produced by the presentation phase
Input:	Report produced by the Presentation phase
Precondition:	Presentation phase completion
Postcondition:	None
Product:	Digital triage assessment

3.1.7 Preservation

Preservation is an overarching requirement for any activity involving digital evidence. During every phase of the process the triage examiner must insure that no change is made. The technology, software and hardware, exists to prevent any writing to the media under examination. If the media is being examined through the use of an examination machine, then a hardware write blocker must be in place between the source media and the examination machine. If the triage examiner is unable to remove the hard drive or a examination machine is not on site, then a tested live boot CD such as Helix(http://www.e-fense.com/products.php) can be utilized to boot the system into a safe environment.

In both cases since the original evidence was accessed, whatever process is used to do the triage will have to be recorded and that record will have to follow the evidence through the traditional forensics process if one is applied.

3.2 Digital Triage Process Model Usage

Digital triage and this process model both have different purposes depending on the location of the media being examined and whether the examination is being conducted on live or dead evidence. Digital evidence is going to be in one of three places the field, the office, or the laboratory. In the field digital triage can be used to provide feedback to the search and seizure team, to gather quick information, and for use in suspect interviews.

A lot of local law enforcement agencies do not have their own digital forensics lab. If a piece of media is located at the law enforcement office waiting to be sent off for examination, it might be wise to first perform triage to determine if it is even worth sending in. If it is determined through the use of this process that it is unlikely that evidence is present, then perhaps the investigators will want to concentrate their search for evidence in other areas and not wait on an exam. The simplicity of this process allows for its use in such a situation where training is limited.

A digital triage process that assists in prioritization would be extremely helpful in increasing efficiency for digital forensics laboratories. For most digital forensic labs, working off a backlog has become the common practice. It is important for digital forensic labs to be able to prioritize cases based on the available evidence, and in some circumstances even to refuse to accept cases until their backlog is reduced. The automation described in this model can make it fast enough to be useful in this type of situation.

3.2.1 Common Digital Triage Scenarios

A commonly mentioned use for digital triage is that of a kidnapping case due to the need for a speedy turn-around. In this situation the primary source of evidence is typically already known, and the digital triage examiner already knows what to look for. The template could however contain keywords to determine what chat programs are installed, and key registry entries that will report passwords for chat programs and most recently accessed files. Knowing this knowledge ahead of time will assist the digital triage examiner in more quickly obtaining the specific information needed in the Triage Examination phase.

Another commonly discussed use of digital triage is that of a soldier in the theater of war. In this type of situation automation and simplicity of use is vital. The solider may have limited training in the use of the tool, and speed is always important to a soldier in combat. One scenario could be a soldier who needs to quickly locate media that was involved in a suspected terrorist activity. Utilizing this model could quickly allow the solider to determine which computer among many contains the primary source of intelligence needed. See Table 3.7.

Table 3.7

Terrorist Activity Template

Example Template	Terrorist Activity
Keyword List	known terrorist groups, known terrorists, bomb parts
File Type Alerts	ArcInfo (mapping software), AutoCad
Know File Alerts	known steganography programs

3.2.2 Digital Triage Trade-offs

The analysis performed in the Computer Profiling and Crime Potential phases calls for the analysis of URLs, a search of keywords, an analysis of file types, and a search for known files. There will need to be some trade-off analysis conducted. The more thorough the searches, the longer the process will take.

One trade-off considered will be whether these keywords should be searched for within all present files, just through the file names, or within certain file types. The hashing of all files to determine if they are a known file will take considerable time as well. Known file searches can be sped up by just searching for known file names. This would not be as accurate as hashing every file, but would provide considerable speed up. In both of these situations, the tradeoff will be between looking at the file names or file contents.

The inclusion of additional processing as briefly mentioned will also be a metric to consider. Additional processes provide additional automated intelligence, but it also provides increased processing time, and more time for triage examiner to decipher the report results. For the actual model testing, a more generalized approach can be taken without additional processing options sticking to the simple standard template described in Table 3.2.

There can be two types of scans that could be performed on a drive during digital triage, a surface scanning looking only at present files in allocated space and a deep scan searching through unallocated and allocated space. Both of these ideas were originally considered for possible trade-off analysis. The idea of doing a deep scan was quickly discarded during initial development. This type of analysis would be too time intensive to be of use in digital triage.

All of these speed versus completeness trade-offs are greatly dependent on situation. For example, if a triage examiner is in the field then speed is probably a very high priority. All searches can be reduced to file name searches, and all known file searches could be limited to searching for the names of the files instead of the hash values. If the triage examiner is back in the laboratory and using digital triage to sort cases, then speed might be less of an issue and the examiner can perform the searches using the file contents as well as the file names. The ability to adjust these parameters will be important to the design of the utility that implements the proposed model.

Some trade-off analysis will be done to determine the correct amount of information to be gathered for experimental purposes. The more information gathered by the automated process the faster the triage process should occur, but the greater the time penalty for running the triage model. It will have to be determined how much time can reasonably be spent during the digital triage process during model testing.

CHAPTER 4

IMPLEMENTING THE AUTOMATION TOOL

Before the model described in this dissertation could be tested, it had to be implemented [9]. This chapter describes the creation of a utility developed to facilitate the testing of the proposed model by implementing its automated phases as shown in grey in Figure 3.1. This utility will be referred to as the FMPU for Fast Modular Previewing Utility. (A manuscript concerning FMPU implementation has been accepted for publication by the Journal of Digital Forensics Security and Law September 2012.) Appendix A contains a sample report from a test run for the FMPU. Samples of all report sections mentioned in this chapter can be viewed there.

4.1 Tool Development

Any intelligence gathering process should be efficient, fast, and as automated as possible to be a successful digital triage process. A series of scripts that gather quick useful information for digital previewing were developed to facilitate the testing of the proposed model. These scripts contain original code, native Linux commands, and open source tools. The goal was to create an automated tool that could quickly profile a computer for use in digital triage. The development and publication of such a utility serves several goals in addition to the testing of the digital triage process model.

1. Determine what information can be gathered quickly from a computer in an automated fashion
2. Add support for the use of digital triage in digital forensics by showing how such a tool can be created
3. Demonstrate a modular framework that can be used by others in the creation of their own tools

The first decision to be made was what information to collect. Some information is universally useful, and the usefulness of other information is dependent on the type of case under examination. There have been calls for research into crime class modeling and research for the gathering of digital forensics corpora that will make this type of modeling possible, but the scientific community still has a long way to go[43, 25]. Therefore, for initial development it was decided to explore what can already be quickly gathered using existing open source tools and scripted using existing commands. Further research in modeling digital triage per class of crime being investigated could assist in refining the utility and defining what additional tools need to be built. The FMPU can serve as a base for such research.

4.1.1 Testing Environment

When selecting a testing and development environment the following requirements were outlined:

1. Digital triage must avoid changing the evidence when examining a dead system
2. It should be possible to perform digital triage in the lab, in the office, or in the field
3. Some form of boot media must be used to incorporate full onsite capability

With these requirements in mind, the Linux distribution CAINE installed to a USB drive was chosen as the development and testing environment. It is already designed as a digital forensics tool and can be installed on a USB drive allowing for the use of read-write media instead of write-once optical media which is also a common distribution method for forensic Linux distributions. Using USB drives for the FMPU vehicle may restrict its use on older machines that do not allow for USB boot. However, transfer from USB boot media to optical media would not be a difficult process. The CAINE Linux distribution was used for development, testing, and for the final vehicle of the utility.

4.1.2 Programming Language and Program Structure

On of the requirements decided on during development of the FMPU was the ability to quickly parse text. With this requirement in mind, Perl was chosen as the language of choice. Perl is well known for its text parsing ability and is already included in the CAINE distribution requiring no modification of the CAINE environment. Perl proved to be a wise selection as it easily facilitated the development of the FMPU.

Another advantage of using a scripting language like Perl is the ease of which it allows the utility to be expanded without the need to recompile the executable allowing a single user or a group to easily add to it. Perl also facilitates the modular design of the FMPU. This modular design is implemented through the following procedure:

1. Main module accepts as input: report name, report location, and module list (optional)
2. Main module writes HTML header
3. Main module gives control over to 1st module on the provided list
4. Module 1 extracts information
5. Module 1 formats information as text, HTML table or separate Web page
6. Module 1 appends HTML table, link, or text to report
7. Main module creates HTML footer to close the report

Steps 3 - 6 are repeated for each module on the provided list or as applicable. For example, if it is determined that the system does not have an operating system installed on it, then certain modules will be skipped such as Windows registry extraction. Expansion of the program is now a simple matter of creating the new module; placing it in the same folder as the other modules; making sure the output is in the form of an HTML table, link, or plain text; and adding its name to the list to be run. RegRipper discussed in the User and Application Information section is designed in a similar modular fashion and also written in Perl.

4.2 FMPU Functionality

Data extraction starts with disk information where it examines the physical and logical disk information as well as file system information. It then looks at the files on the system by examining the quantities of file types that are being stored. Finally, it looks at specific user information recorded by the operating system and applications. Each of these categories will be discussed in detail.

4.2.1 General File System Information Extraction

Taking inspiration from Brian Carrier in his book, “File System Analysis,” the utility was designed to extract file system information at different physical disk and file system levels [10]. General file system information can assist in triage by allowing an investigator to quickly determine how many disks are installed in the system, the size of such disks, and the complexity of the partitioning system on the disks.

At the highest level, a digital triage examiner needs to know what physical disks are on the system and how many partitions are on each disk. This is quickest and easiest of all the information sets to obtain. Most modern Linux systems identify this information during boot up and create mount points in the /dev folder for each physical device and each partition. A simple listing of this folder for mount points starting with “hd” or “sd” will provide the information. A change in the letter following the “hd” or “sd” will indicate the number of physical drives. For example, a system listing sda, sdb, sdc will have three physical devices recognized by the system. There are other ways of acquiring this information, the Linux command “fdisk -l” for example. However, the method just describes produces simple output that can easily be parsed for the final FMPU report.

There should also be a mount point for each partition indicated with a number at the end of the name. For example, a system listing sda1, sda2 would have a physical device referenced with sda that has two partitions labeled 1 and 2. This information can be found by running the command “ls /dev.” There are other things in this folder as well, but filtering the output with another common utility called grep and searching for the letters “sd” or “hd” quickly displays the desired information. The final commands would look something like this:

- ls /dev/ — grep hd.
- ls /dev/ — grep sd.

If a bootable thumb drive is used as a final vessel for the FMPU such as in the testing environment, it will also be necessary to identify which label the thumb drive is given as well to avoid including the thumb drive information in the final output. This option was included in the FMPU prototype to allow for this. However, it might be wiser to avoid having to include this option so as not to confuse the novice digital triage examiner. Providing the tool in an optical media format instead will address the danger.

The sector layout of the disk and the file system label of each volume is determined using the Sleuth Kits *mmls* command as shown in Table 2. This quickly allows the digital triage examiner to locate possible hidden areas on the disk by making note of large amounts of unallocated space, the complexity of the disk layout, the potential for data recovery, and some hint of the expertise of the user. The amount of data recovery possible is dependent on the file system used to store the data. For example, Windows FAT file systems are notorious for leaving data remnants behind, but Linux based file systems are designed in such a way that data remnants are more quickly over-written[10]. Also, a disk with multiple types of file systems or file systems that are less commonly used could indicate a more advanced user that is willing to experiment with different file systems instead of a user who sticks with the file system preinstalled on the media.

Table 4.1

Sample output from Sleuth Kit *mmls* command

	Slot	Start	End	Length	Size	Description
00:	Meta	0000000000	0000000000	0000000001	0512B	Primary Table (0)
01:	—–	0000000000	0000002047	0000002048	0001M	Unallocated
02:	00:00	0000002048	0001257471	0001255424	0613M	NTFS (0x07)
03:	00:01	0001257472	1953523119	1952265648	0930G	NTFS (0x07)
04:	—–	1953523120	1953525167	0000002048	0001M	Unallocated

Essentially from this report, an examiner can see how many partitions are on the system, the file system types, and if there is any space that is not allocated to a partition. The file system type, for those who are more versed in technology, will be an indication of how advanced the user is, the type of information that is on the system, and large amounts of space not allocated to a partition could indicate data hiding. Further research and development can be done to produce additional output indicating red flags whenever there is an above normal amount of unallocated space or when a file system type that is not common is identified. The Sleuth Kit command *fsstat* can provide a lot of additional detail concerning the file system, but the only other item included in the FMPU prototype final report is *fsstats* attempt to guess the operating system on each volume.

4.2.2 File System Metadata Information

File system metadata refers to the reviewing of the system metadata of all files located on the system, not to be mistaken with internal file metadata such as EXIF information in JPEGs or user information in Microsoft Office products. File system metadata can provide clues to what the computer has been used for. A media machine will most likely have a large number of media files. A machine used more for work is more likely to contain more documents and spreadsheets.

The Sleuth Kit suite of tools also includes a tool called sorter that will classify all files on a system. The idea behind sorter was used, but the tool itself proved too time intensive to utilize in the FMPU. The file classification report produced by the FMPU is instead compiled using native Linux commands. The determination of file type can be done in two different ways. All files are in essence binary data and the examination of this binary data can often be used to identify the file. What makes a file useful is the program used to interpret that binary data into something useful to the user. Windows machines use the two or three letters following the last period in a file name, commonly called the file extension, to determine what program to use for this interpretation. These last three letters can also be used to identify the file type.

Using naming conventions for file type identification has a disadvantage. Windows operating systems typically do not use any verification of file type against the file name. This means nothing prohibits the user from renaming a file to any incorrect file extension. There is also the possibility of a glitch in the system resulting in files having an incorrect extension. For these reasons, digital forensics programs typically depend on the first few internal bytes of a file to determine the true file type.

The FMPU divides the output for file classification information by Windows user directory and by entire volumes. In addition, the user can select first byte signature identification or file extension identification for user directories and volumes independently. The choice of one technique over the other is dependent on the time critical nature of the situation.

As explained, a first byte signature analysis is more trustworthy. However each time a file is identified by signature, its first bytes have to be compared to a list of known byte signatures. This list needs to be extensive to be useful.

In the initial tests byte signature analysis performed on user directories added three minutes of processing time on the testing machine. Doing full byte signature analysis on complete volumes with Windows 7 installations increased this time to over twenty minutes due to the number of files that had to be classified. Full byte signature analysis on non-Windows operating system volumes varied by the amount of files stored on the drive. The compromise adopted was to set the tool by default to do full byte analysis when it encounters a user directory, but to perform file extension analysis when it encounters a complete volume. During later testing, this was changed to file extension analysis for both user directory and volume to facilitate faster testing. This is digital triage not a full examination. Therefore, the sacrifice is acceptable.

Time line information was originally considered as another system metadata category to explore, but it experienced the same slow-down issue. Creating a sorted list of file times for specific directories is reasonable, but creating a listing of an entire drive proved too time-consuming. In addition, the report generated could not be summarized leading to the amount of data the digital examiner has to sort through being beyond reasonable for digital triage in the field or in the lab. Therefore even if time were available, creating a timeline of all files on the drive is unreasonable. Just as with everything in a criminal investigation, the type of analysis that is needed is greatly dependent on the situation. Sometimes it is necessary to identify every activity that occurred within a given window of time. With this time window as a limiting factor the amount of data can become more reasonable. This, however, moves from the realm of digital triage into analysis. For the final utility the time line information was excluded, but can be easily added as an additional module utilizing existing tools in non-digital-triage situations.

4.2.3 User and Application Information

User information is simply the user information recorded by the file system or an application usually without the casual user knowing it. For example, the user created username for each user on the system is created by the operating system itself. Application information is the information created for the user during application use. This information is specific to the application and the operating system being examined. For testing purposes, the concentration was on making tools to search for application information on the Windows environment. However since the tool is run from a Linux boot disk, it would not be difficult to add additional modules to look for artifacts more common to the Apple or Linux environments. After developing such a module, one only has to update the tool to execute the application information module specific to the operating system or application on the system under evaluation.

First, research had to be conducted to determine what user specific information should be collected. In Simpson Garfinkels paper, “Forensic Feature Extraction and Cross-Drive Analysis” one of the techniques explored is the identification of the primary user of a computer by creating a histogram of all the email addresses on the system[24]. This type of analysis matches very closely the goals of this project, and the incorporation of this idea was explored during tool development. However, this technique was designed for detailed analysis not triage. For now, it proved to add too much additional time to the analysis. It is hoped, the idea can be incorporated in some form later.

The idea of feature extraction for profiling was adapted in a different form. During file classification the reports are divided by volume and by user directory. At the end of each link is a number representing the number of file types identified at that level. At a glance the digital triage examiner can quickly determine which volume and which user account has had the most use by looking at the number of file types identified. This is not a guaranteed method, but it can provide some indication. The Web links work in a similar fashion. They are divided by user directory, and a number is provided at the end of the link showing how much history or how many domains are listed. See Appendix A for an example report.

Rogers et al. in their work Computer Forensic Field Triage Process Model discuss the gathering of user specific information for field digital triage[33]. Their model did serve as inspiration. However, their work does not suggest automation as this utility does. Their work concentrated more on identifying and modeling the types of information that should be sought. They were not concerned with creating an automated utility. The FMPU does extract similar sets of data as the data described in their work such as Web history and Windows registry information.

On a volume with user directories, usernames for each user are collected by looking at the user directories as listed on the system. This can provide an indication of how many users are on the system and possibly the identity of those users. However, the digital triage examiner must remember that there is no easy way to tell who is actually using which account. It is also important to note, user directories can be placed in non-standard locations eliminating this benefit. However, for more advanced FMPU users the configuration file can be edited to restore this benefit by specifying the user directory location.

The first version of this utility also incorporated two other sets of information as its target: the Windows registry and Internet Explorer history. The Windows registry was chosen because of the wealth of information that can be quickly obtained from its examination. Several works discuss the structure of the Windows registry and the creation of tools to examine it [14, 22]. However instead of creating a tool from scratch, it was decided to incorporate the existing tool RegRipper. RegRipper is a tool maintained and provided free of charge by Harlan Carvey [15]. The FMPU calls the tool to extract information and provides the results in the final report. Registry file information is separated by volume and by each user. The information about the entire system was produced by examining the SOFTWARE, and SYSTEM registry files. Information for each user is gathered by examining each users NTUSER.DAT file.

The data to be extracted was based on what is possible, what is already coded as RegRipper plugins, and what would be the most useful during a digital triage situation. The following was selected as an initial set of information to extract for testing:

Per User Information Extracted:

- Logon name of the user used to verify the user list
- Websites typed directly into a Web browser
- Recently opened documents
- Recently run items from the command line box
- Media Player recently played files
- AOL Instant Messenger information
- Skype communication program settings
- Yahoo Instant Messenger settings
- MSN messenger settings

System Information:

- List of USB devices that have been attached to the system
- Shut down counts and times

Software Information:

- The default browser

This selective extraction is dependent on which plugins are called during execution. The creation of these plugins is an ongoing project supported by the developer. Newly developed plugins can be added to the RegRipper program by adding the plugin files to a specific directory in the RegRipper program directory. The selection of which plugins to run for the FMPU is done through an input file. Plugins can be added or removed to the report through the editing of this simple input file.

The second piece of information retrieved is Web browsing history. In order to facilitate future Web browsing, the default on most Web browsers is to keep a record of what sites a user has visited. Unless a user changes this setting, a Web history is maintained. Thus, internet history can provide a lot of information about the user such as what email accounts they use, what Web searches they have been performing, and other personal habits. There are various Web browsers, and each keeps a history file in its own format and in its own location.

Internet Explorer history was chosen over other browsers for the prototype because it comes standard on Windows computers and is arguably the most popular Web browser in use today (http://marketshare.hitslink.com/). Further expansion of the FMPU should include modules for additional Web browsers. Internet Explorer stores its history in the index.dat file located in the user directory [44]. Reverse engineering its file format is a simple matter as there are plenty of published works that have laid the foundation work [31, 32]. Along with the Web page visited by the user, the index.dat file also contains the time the Web page was visited, whether it was intentional or a redirect and the associated cached item located in the history folder.

The purpose of the FMPU is to selectively collect those items that are of the most interest to the digital examiner and provide the information in a useful fashion. The goal of the FMPU is not to present all the possible data. Therefore, a similar approach to what was done with the file type analysis was also performed for the Web browser history analysis. Instead of listing all the URLs and all the information about the URLs, each domain visited is counted. For example, if a user visits www.Website.com/link2 and then www.Website.com/link1 the FMPU will only record www.Website.com, 2. The final listing is then sorted by number and sent to the output. The raw output used to create this list is also included in the report in case the triage examiner needs more detail.

Presenting the data in this manner allows the triage examiner to quickly determine what Websites have been visited and to what depth or frequency. A single visit to a Website could indicate a redirect or accident. A higher number will indicate multiple recorded visits or a much deeper exploration of the Website. Number does not always signify importance however. A Website visited a single time might be an important clue or the same Website visited multiple times may not record each visit. Thus, lower on the list does not necessarily indicate less importance to the examiner.

4.3 Red Flag Report

The primary purpose behind the creation of the FMPU is to facilitate the testing of the model proposed in this dissertation. The three phases the FMPU will implement are Computer Profile, Crime Potential and Presentation. The idea of the Crime Potential phase is to allow customization based on the class of crime being investigated. Crime class customization is an open area for research, and there are not a lot of available testing sets. With real evidence restricted from civilian access, it will be hard to develop any for future works[43, 25].

The red flag report is generated by filtering using a list of red flag key words chosen before executing the FMPU. All information is filtered by this red flag list and included in a separate report similar to the main report generated by the utility. For example, file names are gathered as part of the file type quantification analysis. These file names are run through the red flag list and any file name that has as part of its name a red flag word will be listed in the red flag report. A similar process is done during the gathering of Website history.

In this way, a digital triage examiner can quickly find information based on crime class. A similar process is carried out with most commercial forensic tools. In a commercial tool this process will typically take one of two forms. Either files will be identified by hash value or keywords will be searched for throughout the disk. Inclusion of either of these options would be too time intensive. The idea to search only the information that is being gathered is an ideal compromise for digital triage. The creation of standard red flag keyword lists would be a great area for further development.

4.3.1 Semi-automated Crime Specific Digital Triage Process Model Implemented

The primary purpose for creating the FMPU was to implement the automated phases of the Semi-automated Crime Specific Digital Triage Process Model. Those automated phases are Computer Profile, Crime Potential, and Presentation. The Computer Profile stage calls for the extraction of logical and physical disk layout, file system information, and application information. As shown above, the FMPU does extract all this information.

The Crime Potential phase is meant to help the examiner focus the triage on the specific class of crime. This is implemented through the use of a red flag filter. This filter watches all data for keywords previously entered by the examiner and moves any data item it finds into an alert report for viewing. The other profiling mentioned in the model introduction was that of significant amounts of a particular file type(s). The FMPU gathers and counts all file types on each volume then sorts the output by number of occurrences. A quick glance at the top of this report will provide this information. The digital triage examiner can determine from their template the type of files that should cause concern and by viewing the report see which volumes has the largest quantities of this file type. There is no current data on what an average user has on their computer as far as file types. Therefore, there is no current automated way to red flag solely based on the number of files on the system. The digital triage examiner will have to use their own template specific to the case under examination.

The Presentation phase is the presentation of the results in a useful manner that can help guide the digital triage examiner in their efforts. The FMPU produces this report in the form of easy to navigate HTML pages. HTML is already commonly used for digital forensic reports and allows for the easy inclusion of the text reports produced by the FMPU. For an example of the FMPU report see Appendix A.

4.4 FMPU Future Work

This first iteration of the FMPU was a prototype with presets chosen by the designer. Possible future upgrades for this utility could include:

- Default settings chosen in a more scientific manner
- Multiple Web browser support
- Option for a deeper file or sector scan
- Live system option

The selection of which file type classification to do, first byte or naming convention, and which registry keys to extract was made based on the developers experience and in-formal interviews with currently working digital forensics examiners. Experiments were carried out with these settings and have the goal of testing the usefulness of the tool and the methodology it helps to implement. Once the value of the tool is verified, further testing should be carried out to better determine what default settings are the most useful. With file type determination there is an element of processing time that has to be evaluated, and with the registry keys there is the consideration of how much information is too much as it contributes to user evaluation time.

The FMPU was designed to find and extract Web browser history for Internet Explorer. There are, of course, other Web browser options available to each user. A more comprehensive scan would include the search for these Web browsers as well. Also, currently the FMPU only extracts the URLs listed in the history. Another question that needs to be explored is the usefulness of including the other information such as time stamps and direct connection versus redirect information. The goal of the FMPU is to stream line all information to facilitate quick digital triage decisions.

For keyword searches, the thoroughness of scans can be divided into 3 levels. The FMPU looks only at the information already being gathered when building its alert report, for example file names and Web history. This could be considered a level 1 scan. A level 2 scan would also include the scan for words inside files. A level 2 scan would take considerably more time. How much time would be dependent on the number of files present on the system. A level 3 scan would be a sector-by-sector search for keywords. Level 3 scans would take the most time, and be time dependent on the drive size itself. Level 1 scanning was chosen based on the idea that digital triage has to be as quick as possible to be useful. This is certainly true when performed in the field. The other two situations already mentioned were, in the office for determining if a full examination is warranted, or in the lab for case prioritization decisions. These situations are not as time critical and what level of scan that would be the most useful would be an interesting area for future work as well.

Digital triage can be performed on either live or dead evidence. Live evidence would be considered an operating system that is currently running, and dead is one where the system is controlled by a digital forensic entity such as a separate analysis machine or an operating system booted from external media. Currently the FMPU only functions on dead media having been developed from the Linux bootable distribution CAINE. The analysis techniques used by the FMPU are applicable to a live system, but will not currently function in a Windows environment. Also, any live analysis tool should be run in as small of a footprint as possible. This was not a concern during the creation of the FMPU. Future work could be to create a live port of this tool.

CHAPTER 5

EXPERIMENTAL DESIGN

Chapter 3 introduced the Semi-automated Crime Specific Digital Triage Process Model[9]. Chapter 4 discussed how the model was realized through the creation of the FMPU. This chapter describes the experimental design for the proposed hypothesis, the statistical analysis of the results, and presents the results of testing. Final conclusions are presented in chapter 6.

Mississippi State Institutional Review Board approval was received before any experiments were performed. Dixie State College Institutional Review Board board approval was obtained prior to performing experiments at the Southwest Regional Cyber Crime Institute.

5.1 Testing Plan

The hypothesis proposed in this dissertation is that the speed it takes to make an assessment during digital triage can be decreased by at least 50% with the use of the digital triage process introduced in this dissertation, and the digital triage process proposed will not decrease the accuracy rate of the decisions made during digital triage.

After dividing the testing subjects into separate trials, it was also decided to address whether or not expertise had an effect on time and accuracy as the groups naturally fell into categories of expertise while assigning subjects to trials. See Table 5.1 for a summary of experiments.

Table 5.1

Experiment Summary

Hypothesis	The use of the proposed model during digital triage will decrease the time it takes to perform digital triage, as digital triage is defined in this proposal, by at least 50% without decreasing the accuracy of decisions made during digital triage.
Independent Variables	The report produced by FMPU and subject expertise
Dependent Variables	The accuracy and speed of the decision made by the subject
Constants	The digital media sets used to make a decision and the case scenarios
Groups	Testing group consisting of an experimental group and control group; validation group consisting of subjects who are qualified examiners
Statistical Method	Multivariate Analysis of Variance (MANOVA)

5.1.1 Testing Group Primary Separation

There were two primary groups of test subjects, a testing group and a validation group. The validation group was a small group of individuals used for validating the tool on real evidence and consisted of fully qualified examiners. (Fully qualified examiner was defined as an individual who has at least one professional certification in digital forensics, and who is actively performing digital forensics examinations as part of their job.) The testing group was comprised of college students actively taking digital forensics courses. The validation group provided for qualitative testing of the FMPU, and the testing group provided a quantitative analysis. The testing group was further divided into experimental and control groups. Each subject was assigned to 1 of 5 trial groups based on testing location and expertise. The testing group had some knowledge of the computer forensic process, but were not active examiners.

Besides the obvious benefits of having validation testing, the need for a validation pool is due to the restriction of accessing real case data, the lack of forensics test sets available to academia, and the availability of qualified examiners to use for testing. The data sets used for testing were manufactured. The testing group was used to test the hypothesis and the smaller validation group was used for validation of the tool.

It was stated earlier that digital triage would be useful in the office to determine if evidence should be sent to a lab for examination, in the field to help with search and seizure efforts, or in the digital forensics lab for case prioritization. In the law enforcement office the digital triage user is likely to have little to no digital forensics training. In the field your digital triage examiner is likely to be part law enforcement and part examiner with an intermediate level of training. Finally, your examiner in the lab is likely to be an expert. Therefore although not law enforcement, college students in a digital forensics program still serve as credible subjects. Digital triage is a process that is carried out at all levels of expertise, and the FMPU was intended to be useful to all expertise levels.

5.1.2 Testing Environment

Validation subjects used their own chosen test set, and testing was performed in a digital forensics lab. The monitor had little to no interaction with the actual evidence due to legal concerns. The evidence used for the testing group was generated test sets. Test drives were generated by two different student volunteers not directly associated with the experiment and with minimal knowledge of what the drive was to be used for, and these individuals had minimal knowledge of digital forensics and little or no knowledge of digital triage.

Each volunteer was provided a computer with the same base image. This base image was of a digital forensics training computer that had been in service for 2 semesters without being reset. The base image computer was selected randomly. Usernames for the computer were generic to avoid any misleading information. The subject was then asked to use the drive for normal activity as well as to use it to plan/conduct a fictitious crime. Normal activity was explained as activities normally done on a computer in a casual or work day. Volunteers were also asked to avoid anything too personal like accessing their bank accounts or personal social networking sites.

Each volunteer was asked to perform these activities for approximately 8 hours. On completion they created a short scenario describing their fictitious crime. The computer used by the volunteer was then imaged to preserve it for later testing. The evidence creation computer was re-imaged to the base image before the next volunteer used it. The two fictitious crime cases created were a murder scenario and a child pornography collector (substituting images and phrases of kittens for actual children).

The actual crime scenarios provided to each test subject during testing were simplified versions of the scenarios turned in by the volunteers. The kitten pornography scenario was described as an individual who collects, trades, and downloads pictures of kittens. The murder scenario was described as a man who's wife had gone missing, and this same man was recently picked up in the local forest with a shovel. This information was provided to each test subject at the beginning of testing.

The testing environment was set up with each test machine identical by imaging them with the base set used to represent no criminal activity. Two external drives were attached to each test machine, one imaged with the murder scenario and the other with the kitten pornography collector scenario. The FMPU report is divided by drive designation allowing for a single run on all three drives at once during testing. This more easily facilitated the experiments.

Each subject in the experimental group created one keyword list that would encompass both crime scenarios and then ran the FMPU. The control group searched through the drive using the file browser built in to the CAINE environment or any other tool they already knew, remembered from class, or discovered during the experiment. Subjects were allowed to ask simple technical questions during testing, but no information was provided to them about the case or how to interpret any results.

5.2 Validation Group Testing - Qualitative

The following presents the testing procedure and results for the validation testing group.

5.2.1 Validation Group Experimental Procedure

The validation group consisted of 4 fully qualified digital forensic examiners. These subjects were provided the FMPU and a minimal set of instructions on how to use it. Prior to running the tool on their chosen evidence, they were given a demonstration of the FMPU on a test drive, and an explanation of the results. They were then asked to run the tool on a case they had previously examined and could still remember clearly. The experimental procedure was as follows:

1. Subject was asked to give a brief description of the case they intend to use to the tool on.
2. Subject was asked to think back to the case and make a list of 5 things that would have been helpful to know. (These 5 things should be items that would have facilitated digital triage, helped them prioritize evidence, or would have helped guide them during their original exam.)
3. Subject was asked to make a keyword list based on those five things.
4. Subject then ran the utility based on the keyword list and viewed the report.

After the report was generated subject was asked the following questions:

- Of the 5 pieces of information that would have been useful for you to know, how many did you find in the report? Where did you find them?
- Were there items of interest on the report that you had not listed? List them.
- What information would have been useful to include that was missing?

5.2.2 Validation Testing Results

The results of the validation testing were in support of the FMPU. Table 5.2 shows the number of items found versus number of items listed. As shown, all groups found at least 1 of the items predicted after hearing a description of the program. Also, all groups found several items they had predicted, but would have been beneficial. These metrics are qualitative in nature not quantitative. Each subject used their own case, and there was no real control. The value in the numbers shown here, and in the other tables in this section, is that the tool did function on real evidence sets. For more details, see the separate validation test descriptions at the end of this section.

Table 5.3 shows the general opinion of the tool by the test subjects. As shown, the majority of the subjects believed this tool would decrease triage time, exam time, and allow for case prioritization. No validation subjects disagreed or believed the tool would not apply to their situation.

Table 5.2

Validation Test Results

Subject	Location	Case Type	Items Listed	Items Found
1	CCFC	Child Pornography	5	4
2	CCFC	Child Pornography	4	3
3	SWRCCI	Slander	4	1
4	SWRCCI	Child Pornography	4	3

Validation tests 1 and 2 were performed at the Cyber Crime Fusion Center (CFCC) in Jackson, Mississippi. This facility is a multi-department task force actively performing

Table 5.3

Validation Responses

Rating:	Agree	Somewhat Agree	Don't Agree	N/A
Would Decrease Triage Time	2	2	0	0
Would Decrease Exam Time	3	1	0	0
Allow for Case Prioritization	4	0	0	0

digital forensics investigations. Validation tests 3 and 4 were performed at the Southwest Regional Computer Crime Institute(SWRCCI) in Saint George, Utah. This facility is an active digital forensic laboratory established through grant funding at Dixie State College of Utah. The results of each of these tests will now be summarized.

Validation test 1 was performed on a child pornography case involving peer-to-peer downloading and trading. Subject hoped to find the following items:

1. Evidence of peer-to-peer traffic
2. Large amounts of graphics files
3. Internet history and search results related to the crime in question
4. Attached USB devices that could have been used for storage
5. Child pornography terminology

Subject 1 found evidence of peer-to-peer programs on 2 different user accounts, files named with child pornography terms, USB history, and Internet activity related to child pornography. Additionally they found four other items of interest including child pornography related shortcuts, and pertinent registry information. They commented that it would have been beneficial to have the ability to search through document files and view thumbnails. Additionally subject 1 stated that the tool would have quickly narrowed the investigation scope, and helped to focus the examination on the specific users of interest.

Subject 2 also chose a child pornography case. Subject hoped to find the following items:

1. The breakdown of image files on the disk
2. Evidence of peer-to-peer sharing program
3. If the suspect images were from Websites or peer-to-peer sharing programs
4. The number of users on the system

Subject 2 quickly identified all items listed except for whether the primary source of images were from Websites or peer-to-peer sharing. In addition, they found 3 items not listed that would have been useful including the registry information of recent docs, typed URLs, and Web browser history. Subject 2 also commented that the tool would have narrowed down the scope of the investigation, and would liked to have seen some timeline information included to add perspective.

Subject 3 chose a corporate case involving slander and insubordination. Subject 3 hoped to find:

1. Evidence of specific documents
2. Evidence of slander of a specific group of individuals
3. Evidence of slander concerning a specific corporate program

Of those items listed the only item found was evidence of the specific document that held the slander used in the corporate action already taken against this individual. No other items were identified. However, subject 3 did not find evidence of the other items listed during the original investigation either. Subject 3 added the comment that the most beneficial aspect of the FMPU was the amount of time it would have saved during the original examination. Subject 3 felt because it revealed the lack of evidence early on, it would have reduced processing time from 6 hours to 1 hour if it had been used prior to the original examination.

Subject 4 chose a child pornography case involving child seduction through webcam chat. Subject hoped to find:

1. Programs related to webcam chats
2. Regular visits to webcam chat sites
3. Use of a data destruction program
4. Large numbers of jpegs and/or video files
5. Regularly visited websites

Subject 4 was able to see the file type distribution desired, the websites, and some evidence of webcam use, but the remainder of the results was limited. The user had apparently run a data destruction program (CCleaner) on the computer recently, and it appears the user also deleted the Web history as well. The FMPU is currently limited to Internet Explorer. Therefore, the monitor requested that the subject confirm the presence of Internet Explorer. Subject 4 did confirm the default Web browser on their evidence was Internet Explorer, and that some Web history was present just not significant amounts or any significant links in the red flag report.

Subject 4 made the comment that although the information was limited, the tool was still useful as it would have shown this lack of evidence pre-examination. Subject 4 also stated that the largest advantage it would have provided would have been in prioritizing the case. All evidence previously found on this case was carved evidence. Subject 4 felt that running the FMPU prior to the previous examination would have predicted this fact.

On viewing all the results it was felt that at least two upgrades to the FMPU should be considered. First, the ability for the FMPU to list all installed programs. Specific programs can be searched for with the keyword lists, but an examiner can not always predict everything they want to see. A listing of the programs would, among other things, allow a quick check for secure erasure programs, and specific peer-to-peer sharing programs. Further experimentation will have to be done to determine if this program listing would be short enough to be useful during digital triage or what filtering could be applied to make it so.

The second improvement that should, perhaps, be made is limited searches within specific file types. In a previous section keyword searches were divided into level 1 file name only, level 2 file name and file content, and level 3 raw sector search. Perhaps a sub level of 2 could be used to search only within certain files such as documents or HTML files. For example, subject 3 felt the digital triage examination would have provided more conclusive results if documents could have been scanned for keywords associated with the slander and insubordination. The idea of a crime template is part of the Semi-automated Crime Specific Digital Triage Process model. The template used for these tests was the simple template of keywords and file types. The author believes that this testing shows support for the development of more advanced templates for testing.

The FMPU is currently being used by the digital forensics lab at the Southwest Regional Cyber Crime Institute at Dixie State College. These improvements and others are being tested, and further development is taking place from feedback provided as the FMPU is being utilized on real case evidence.

Validation testing sought to support this research by validating the FMPU on real evidence. As already shown, the FMPU does work on real world evidence. Every subject found items anticipated and items that had not been listed before running the FMPU. Even on the tests that were limited in their results, test 3 and 4, it was noted that there was very little found in the original examination either. In tests 3 and 4 the items that were found during the original investigation were also identified by the tool. Several of the test subjects even requested a copy of the tool for real world use. Subject response to the general qualitative questions were also positive as shown in Table 5.3.

5.3 Testing Group Testing - Quantitative

The following explains the quantitative testing of the hypothesis. The testing procedure will be discussed, the trial classification outlined, the statistics used for assessment explained, and the results provided.

5.3.1 Testing Group Experimental Procedure

The testing group tests were quantitative in nature. Test subjects were students with basic knowledge of the digital forensic process, but subjects were not active examiners. Both the experimental and the control group were given a lecture or lectures on digital triage. Subjects performed short exercises or viewed demonstrations associated with each topic as part of the lectures. These lectures covered the following topics:

- Digital triage and previewing basics
- Linux boot environments both CD and USB
- Mounting and viewing attached disks in Linux
- Finding, listing, and sorting files in Linux
- Using the strings command to strip and view file contents
- Using the RegRipper utility
- Using the File command

These concepts are the core of the FMPU, and covering these topics will ensure that most subjects can perform digital triage at a basic level before beginning the experiment. Students were encouraged to take notes and write down the individual commands. They were allowed to bring these and any other materials desired to testing. Test subjects were further divided into 5 trials based on level of expertise and location of the testing.

Each test subject was asked to classify each drive attached to the test machine by crime category. The experimental group used the FMPU and the control group used the more traditional ad hoc procedure. The classification of these drives by crime category represents the digital triage examiner's ability to make an assessment about digital evidence. Timing began at the point the examination process began. The hypothesis will be deemed correct if the experimental group takes at least 50% less time on average than the control group to make this assessment without a decrease in accuracy. Subjects were also asked the confidence level of their selection either very confident, somewhat confident, or complete guess. This will allow for outlier evaluation based on those who gave up or just guessed at the answers.

The experimental procedure for the experimental group was as follows:

1. Subject was given a brief description of the FMPU including how it works, and the types of information it provides.
2. Subject was provided a computer with three external drives.
3. Monitor walked subject through running the FMPU and viewing the report.
4. Subject was provided two brief crime descriptions.
5. Subject was asked to make a keyword lists for the different crime types.
6. Subject then made a run of the utility with the keyword list.
7. Subject was then asked to label each drive A, B, and C. A - Evidence of child(kitten) pornography, B - Evidence of murder preparation, and C - Nothing of evidentiary value

The control group procedure was as follows:

1. Subject was provided a computer with three external drives..
2. Subject was provided two brief crime descriptions.
3. Subject was then asked to label each drive A, B, and C. A - Evidence of child(kitten) pornography, B - Evidence of murder preparation, and C - Nothing of evidentiary value

5.3.2 Experimental Trial Classification

Subjects were divided into 5 trials based on expertise and testing location. These trials will be briefly described. Trials 1 and 2 took place at the Cyber Crime Fusion Center in Jackson Mississippi. Test subjects for Trials 1 and 2 consisted of law enforcement officers participating in free training from the Mississippi State University National Forensics Training Center. Instruction was provided as part of their training and participation in the experiment was offered on completion of the training. Trial 1 consisted of 4 subjects and trial 2 had 8 subjects. Subjects expertise for these first two trials was labeled Intermediate. Subjects had not had a great deal of digital forensics training, but what they lacked in digital forensic training they made up for in real world investigation experience.

Trials 3-5 all took place at Dixie State College of Utah. Subjects were given the same lectures and demonstrations as trial 1 and 2 subjects, but testing was carried out during scheduled time slots over a 7 day period. All subjects were students currently enrolled in digital forensic courses. Dixie State College offers a series of courses in digital forensics that lead to either a criminal justice degree with a digital forensics emphasis or a certificate of completion in digital forensics.

Trial subjects expertise level for trials 3-5 was determined based on their current progress through their chosen college program. Trial 3 subjects were enrolled in 3000 and 4000 level classes and were classified as Advanced. Trial 4 subjects were enrolled in 2000 level courses and were classified as Intermediate. Lastly, trial 5 students were enrolled in 1000 level courses and were classified as Novice. See Table 5.4 for a summary of the trial classifications.

Table 5.4

Trial Classification Summary

Trial	Group Size	Location	Description	Expertise
1	4	CCFC	Police officers mixed experience	Intermediate
2	8	CCFC	Police officers mixed experience	Intermediate
3	15	SWRCCI	3000 - 4000 level students	Expert
4	12	SWRCCI	2000 level students	Intermediate
5	11	SWRCCI	1000 level students	Novice

5.3.3 Experimental Procedure Evaluation Trial 1

It should be noted that although trial 1 is included in the results it was not used in the computation of the total final mean times. It was, however, used in the final computation of accuracy. In addition to the primary goals of this research, trial one was conducted in an effort to test the effectiveness of the data sets, evaluate the use of the tools on three data sets at once, and to further refine the presentation materials presented to subjects prior to testing. Trial 1 subjects consisted of experienced law enforcement officers with a variety of computer and digital forensics skills. These subjects were part of an advanced digital forensics training course offered by the National Forensics Training Center at Mississippi State University. Subjects were offered a training session on digital triage and the digital triage environment used by the FMPU. Students were then divided randomly into a control group and an experimental group.

There was not much variation between the times of experimental and control subjects for trial 1, and there was only a decrease of 19% in average times in favor of the experimental group. There was, however, an increase in average accuracy from 50% to 67% in favor of the experimental group. (Data extracted from Table 5.9 Summarized Results by Trial shown at the end of the chapter.) One of the contributing factors for this can be attributed to the small size of the trial 1 test population. It was questioned as to whether the tool itself was also a contributing factor.

The test procedure and tool function were closely examined after trial 1 was completed. The computer used for testing connects all 3 test sets at the same time to make testing easier. This, however, creates a large amount of wait time before the subject begins the examination process. The FMPU examines each volume in turn and produces the report separated by physical disk, logical volume, and user when possible. Each of the three test sets takes approximately 4.5 minutes to process leading to over 15 minutes of collective processing time total. Whereas 4.5 minutes does not seem like a lot of time, the 15 minutes it takes to perform a combined analysis may stretch a test subjects patience for the experiment and more importantly seems to take too much time to provide an advantage over the ad-hoc procedure used by the control group.

On analysis of the FMPU it was noted that 80% of the processing time is taken in the file classification subroutine. In an effort to determine what the volume is used for the FMPU classifies all of the files on a volume by file type, and creates a summary of this information. The first iteration of the FMPU performed this identification based on the first few bytes of the file, a common technique used by digital forensics tools. This analysis requires the comparing of each files first few bytes against a list of known first byte signatures. An alternate file identification technique is using the last 2 or 3 letters at the end of the file name, commonly called the file extension. This type of identification is less accurate as a user or an application can intentionally misname the file in an attempt to obscure data. However, for the purposes of digital triage this is acceptable. Digital triage is about collecting quick intelligence not performing analysis.

It was also observed during trial 1 that the file classification did not seem as vital a component as the more specific application information gathered. With this in mind prior to trial 2, the tool was set to perform file classification based on file extension instead of first byte signature. This change decreased the total processing time on three drives from 15 minutes to 4 minutes. The remaining trials were conducted with file extension identification not first byte signature identification. Which type of classification employed can be easily chosen with the FMPU configuration file.

Trial 1 did, however, provide validation of the quality of the test sets. Two users were able to identify all test sets correctly with at least some confidence. It was also observed that even with the delay caused by the file identification the crime classification could be done within a reasonable amount of time for testing. Results also showed a wide array of accuracy levels. Some subjects were completely correct, some were partially correct, and a few totally wrong. These observations were further supported by the the remaining trials.

5.3.4 Analysis of the Testing Group Experiments

The independent variables for this research are the presence of the FMPU report and level of expertise. Dependent variables are time and accuracy. Complete control group results are presented in Table 5.7, and complete experimental group results are presented in Table 5.8. Individual trial results are presented in Table 5.9. These three tables may be found at the end of this chapter.

The most apparent result is that of the change in accuracy between the experimental and control groups. As shown in Table 5.9 all but two of the experimental group subjects were able to correctly identify each drive. Whereas, Table 5.8 shows the correctness varied considerably within the control group. In the experimental group not everyone felt totally confident of their assessment but, no subject felt their selection was a complete guess.

The evaluation of time is more complex. It is not unreasonable to think that the FMPU report would provide a larger benefit for novices than intermediate users and a larger benefit for intermediate users than for experts. Thus, subjects were divided into trials based on digital forensics expertise and location of testing. This division was in part an effort to explore this predicted effect. However, as shown in Table 5.9 the results did not support this prediction. Trial 3, classified as experts, had a mean time decrease of 71%. Whereas trial 5, classified as novices, had a decrease of only 43% in mean time.

It could also be assumed that a higher level of expertise in digital forensics would facilitate quicker and easier use of the tool. Average time for experimental group subjects support this assumption as can be seen in Table 5.5. Mean time to complete the experiment decreases with level of expertise. However also shown in Table 5.5, this does not hold true for the control group subjects.

It is suspected that the lack of predictable responses in the control group is caused by a higher degree of random chance and level of expertise with the Linux environment in which testing occurred. Control subjects with more experience in the Linux environment were possibly more confident in their results, but likelihood of success and speed was dependent on if and how quickly they found data that allowed them to make a decision. After all, with an ad hoc approach finding the pertinent files can be a matter of chance. Another element of randomness was introduced by those subjects who guessed and caused a skew in the results. Subject 2-3 took 45 minutes before making a complete guess, but subject 4-6 gave up in 13 minutes. Randomness does not invalidate the results, but it does make the analysis more complex.

Table 5.5

Comparison of Mean Time and Accuracy Grouped by Expertise

Group	Expertise	Average Time(m)	Standard Deviation	*Accuracy
Experimental	Novice	19.00	14.85	100%
Experimental	Intermediate	16.6	7.09	100%
Experimental	Experts	10.89	9.77	93%
Control	Novice	33.33	20.59	44%
Control	Intermediate	32.00	18.09	40%
Control	Experts	37.14	19.55	70%

*Percent correct out of 3 possible

The effect of the classified expertise on accuracy and mean time was further investigated. It is suspected that the expertise classification had little or no significant effect on time or accuracy of the group as a whole. The experimental group all followed the same process with assistance from the automated FMPU normalizing their time, but the control group subjects each approached the problem in an ad hoc fashion leading to chance playing a larger role in the outcome. This leads the author to believe the best evaluation metrics is that of the entire group combined or limiting the comparison to those subjects who correctly classified all three drives to reduce the effect of those outliers caused by chance guessing and giving up at random times.

The multivariate analysis of variance(MANOVA) test was chosen to test the interaction between the independent variables(expertise and FMPU report presence) and the dependent variables (time and accuracy of selection). A MANOVA test compares the means of several groups to determine the statistical significants of those groups. It addresses the interaction significance between the dependent and independent variables. MANOVA is also typically used when there is a fear of noise caused by the interaction of the variables. In this circumstance noise can be attributed to random chance in the control group and varied expertise in digital forensics or in Linux. MANOVA reports the different interactions of the variables as separate univariate results as part of the primary multivariate analysis[29].

The MANOVA univariate tests are provided in Table 5.6. As shown the effect of expertise on time resulted in an F=0.131 and a P=0.878 and so the test fails to reject the null hypothesis that there is no difference between the expertise groups for time. The effect of expertise on accuracy resulted in an F= 0.863 and a P=0.429 and so this test also fails to reject the null hypothesis that there is no difference between the expertise groups for accuracy. This shows that expertise has no statistical significance on the test results.

The MANOVA further reported the effect of the presence or absence of the FMPU report on time with an F=18.40 and a P=0.000. The presence or absence of the FMPU report effect on accuracy was reported as an F=18.99 and a P=0.000. The Wilk's Lambda for the multivariate test itself was report as F=0.462 and p=0.000 supporting the validity of multivariate results. In reference to both time and accuracy this rejects the null hypothesis that the groups are the same showing the means are significant between the control and experimental groups.

Table 5.6

MANOVA Univariate Results

Independent Variable	Dependent Variable	Degrees of Freedom	F	p
Expertise	Time	2, 44	0.131	0.878
Expertise	Accuracy	2, 44	0.863	0.429
FMPU Report	Time	1, 44	18.40	0.000
FMPU Report	Accuracy	1, 44	18.99	0.000

Based on the above statical evaluation the metric that best describes the results of the experiment is that of total experimental mean compared to the control group mean. For additional analysis the results of only the groups who got everything correct are of interest as well. All these results, the final conclusions, and future work can be found in Chapter 6.

Table 5.7

Control Results

Subject ID	Time (minutes)	Number Correct *	Level of Confidence **
1-1	42	0	Complete Guess
1-2	36	3	Totally Confident
2-1	21	1	Somewhat Confident
2-2	31	0	Somewhat Confident
2-3	45	1	Complete Guess
2-4	33	0	Somewhat Confident
3-1	38	3	Totally Confident
3-2	12	3	Totally Confident
3-3	20	1	Totally Confident
3-4	38	1	Somewhat Confident
3-5	54	1	Somewhat Confident
3-6	69	3	Totally Confident
3-7	29	3	Somewhat Confident
4-1	66	3	Somewhat Confident
4-2	55	3	Somewhat Confident
4-3	13	0	Somewhat Confident
4-4	18	3	Totally Confident
4-5	25	0	Complete Guess
4-6	13	1	Complete Guess
5-1	12	1	Somewhat Confident
5-2	34	3	Complete Guess
5-3	20	0	Somewhat Confident
5-4	19	0	Complete Guess
5-5	65	1	Somewhat Confident
5-6	50	3	Somewhat Confident

* Out of 3

** Options are : a. Totally Confident, b. Somewhat Confident, c. Complete Guess

Table 5.8

Experimental Results

Subject ID	Time (minutes)	Number Correct *	Level of Confidence **
1-1	42	3	Somewhat Confident
1-2	21	1	Totally Confident
2-1	14	3	Somewhat Confident
2-3	11	3	Somewhat Confident
2-4	10	3	Totally Confident
3-1	9	1	Somewhat Confident
3-2	12	3	Somewhat Confident
3-3	13	3	Somewhat Confident
3-4	7	3	Somewhat Confident
3-5	12	3	Somewhat Confident
3-6	10	3	Totally Confident
3-7	9	3	Totally Confident
3-8	13	3	Totally Confident
3-9	13	3	Totally Confident
4-1	13	3	Totally Confident
4-2	18	3	Totally Confident
4-3	28	3	Totally Confident
4-4	18	3	Totally Confident
4-5	28	3	Totally Confident
4-6	10	3	Totally Confident
5-1	33	3	Totally Confident
5-2	12	3	Somewhat Confident
5-3	12	3	Somewhat Confident
5-4	11	3	Totally Confident
5-5	27	3	Totally Confident

* Out of 3

** Options are : a. Totally Confident, b. Somewhat Confident, c. Complete Guess

Note: Test subject 2-2 had a hardware failure during the test.

Table 5.9

Summarized Results by Trial

Trial	Group	Mean Time(minutes)	StandardDeviation	*Mean Accuracy
1	Control Group	39.00	4.24	50%
1	Experimental	31.50	14.85	67%
Percent Decrease:	19%			
2	Control Group	32.50	18.42	51%
2	Experimental	11.67	7.13	97%
Percent Decrease:	64%			
3	Control Group	37.14	19.55	71%
3	Experimental	10.89	9.77	93%
Percent Decrease:	71%			
4	Control Group	31.67	23.03	56%
4	Experimental	19.17	7.49	100%
Percent Decrease:	39%			
5	Control Group	33.33	20.59	44%
5	Experimental	19.00	10.27	100%
Percent Decrease:	43%			
All Correct	Control Group	40.70	19.10	100%
All Correct	Experimental	16.35	8.97	100%
Percent Decrease:	60%			
All	Control Group	33.91	18.42	51%
All	Experimental	14.91	7.13	95%
Percent Decrease:	53%			

*Percent correct out of 3 possible

CHAPTER 6

CONCLUSION AND FUTURE WORK

With the means between the experimental group and control group confirmed statistically different by the MANOVA test, the final results can now be considered. However as shown, the 5 trial separation by expertise classification was shown not significant. Therefore, no post hoc analysis was conducted with individual trials.

6.1 Main Hypothesis and Research Question Results

The main hypothesis of this dissertation is that the Semi-automated Digital Triage Process Model could reduce digital triage assessment by at least 50% without decreasing accuracy. This model was realized through the creation of the Fast Modular Profiling Utility (FMPU). Testing was designed around the FMPU to research this hypothesis. Testing subjects were organized into 5 trials based on testing location and digital forensics expertise leading to two dependent variables, the presence or absence of the FMPU and level of digital forensics expertise.

The level of expertise, as classified by the trials, was deemed not significant. Digital forensics expertise had little effect, but it is further believed that Linux expertise might have had some effect. Further statistical analysis showed that the presence or absence of the FMPU had a statistically significant effect on time and accuracy. This leads to the conclusion that the Semi-automated Digital Triage Process Model implemented with the FMPU is useful in digital triage regardless of the digital forensics expertise of the user.

Even though the effect of expertise was shown insignificant on the group as a whole, it did seem to have some effect on the experimental group who were utilizing the FMPU. Average time decreased from 19.00 minutes to 16.6 minutes to 10.89 minutes from the novice, intermediate, and expert respectively, but the level of accuracy was rated 7% lower for experts when compared to novices and intermediate users. (See Chapter 5 Table 5.5.) If we assume expertise did have an effect on the experimental group as it seems to, then novices actually seem to perform better with the FMPU than experts. If we assume expertise did not have an effect, then novices did just as well as experts.

This lack of need for digital forensics expertise could be useful to the digital forensics community. Take for example a laboratory environment. Reducing the backlog of digital forensics laboratories was previously mentioned as one of the benefits of employing digital triage. If digital expertise is irrelevant to the accuracy of digital triage assessment when using the FMPU, then it could be used by novices in the digital examination laboratories to triage cases. Novices in this instance could be new employees, outside law enforcement, or investigators who do not normally do examinations. This could decrease case backlogs and increase laboratory efficiency by allowing for case prioritization, allowing for case dismissal before being checked in to evidence, and quicker intelligence for situations where the digital evidence is not likely to go to court such as suicides or plea bargaining.

With the significance of digital triage expertise shown not significant to the group as a whole, the most valuable metric is that of the entire control group compared to the entire experimental group. The decrease in time from experimental to control when comparing all subjects was 53%. An additional test to compare just those with 100% accuracy, in an effort de-emphasis the results from those who made their selection randomly by guessing, resulted in a 60% decrease in time. Both of these values support the hypothesis.

The second part of the hypothesis was that the proposed model would not decrease accuracy. Comparing the groups as a whole the increase in accuracy was from 1.52 out of 3.00 for the control group to 2.84 out of 3.00 for the experimental group. Accuracy was not only not decreased, but it was increased by 54%.

6.2 Research Weaknesses and Future Work

There is not a lot of research in the area of digital forensics and digital triage. Therefore, this research provides several avenues for future work. This section mentions a few directions that work could take using this research as a base. For example, the author acknowledges that the size of the testing pool was relatively small, and future testing could be done to include more subjects for testing. Other future works could include alternate implementations of the proposed model, different user classifications, and further crime template development.

Testing was done through one implementation of the model. A different implementation could produce different results. However, this single implementation does support the value of the model itself. Included in these alternate implementations could be any or all of the upgrades mentioned in this manuscript. For example, limited file searches within specific file types could be implemented and tested.

Although digital triage is undertaken by users at all levels of expertise, the use of more experienced examiners could also show different results. However, this tool certainly demonstrates that the proposed model is useful for users that are not active examiners.

Originally the classification of expertise was based on digital forensics expertise. The model should be tested by levels of Linux expertise as well. As no data was kept about the test subjects, there is no way to go back and re-analyze the data based on Linux expertise classification. Future work could be to re-test with users classified based on this expertise instead and observe the results or re-run the experiments with users all at the same level of Linux expertise to remove it as possible noise.

The Semi-automated Digital Triage Process Model calls for crime class customization based on a template. The basic template used in this testing was a keyword template to red flag results of interest, and the examination of what file types are on the system. Crime class modeling is an open area for research and incorporating different templates specific to different crime classes would be useful research pursued from this research.

6.3 Final Remarks

In this dissertation a model was proposed and peer reviewed through publication[9]. It was then implemented and that implementation was also peer reviewed and currently pending publication with the Journal of Digital Forensics Security and Law. The final results have also been submitted for peer review. Through a validation and testing group the proposed model, implemented through the FMPU, was tested. Results show that it does indeed decrease digital triage time by at least 50% and that it does not decrease accuracy.

Contributions of this work include a new tested digital triage process for the scientific community and a new digital triage tool. It is also the hopes of the author that the publication of such a tool will encourage others in the creation of their own utility. Further contributions of this research include at least 2 articles added to the body of scientific knowledge, and additional support for digital triage modeling within the digital forensic practicing community.

REFERENCES

[1] "National Software Reference Library," June 2011.

[2] M. Anderson, "NTI White Paper: Hard disk drives – Bigger is not better, increasing storage capacities, the computer forensics dilemma," March 2011.

[3] V. Baryamureeba and F. Tushabe, "The Enhanced Digital Investigation Process Model," *Digital Forensics Reseach Workshop*, Baltimore, MD., 2004.

[4] N. Beebe and J. Clark, "A Hierarchical, Objectives-Based Framework for the Digital Investigations Process," *Digital Forensics Reseach Workshop*, Baltimore, MD., 2004.

[5] G. Bell and R. Boddington, "Solid State Drives: The Beginning of the End for Current Practice in Digital Forensic Recovery?," *The Journal of Digital Forensics Security and Law*, vol. 5, no. 3, 2010.

[6] A. Bogen and D. Dampier, "Structured Forensics Examination Planning with Domain Modeling: A Report of Three Experiment Trials," *Journal of Digital Forensic Practice*, vol. 3, no. 1, 2010, pp. 23–32.

[7] A. E. Brill, M. Pollitt, and C. M. Whitcomb, "The Evolution of Computer Forensic Best Practices: An Update on Programs and Publications," *Journal of Digital Forensic Practice*, vol. 1, no. 1, 2006, pp. 3–11.

[8] C. L. Brown, *Computer Evidence: Collection and Preservation*, Charles River Media, Inc., 2009.

[9] G. Cantrell, D. Dampier, Y. Dandass, Y. Niu, and C. Bogen, "Research Toward a Partially-automated, and Crime Specific Digital Triage Process Model," *Computer and Information Science*, vol. 5, no. 2, 2012, pp. 29–38.

[10] B. Carrier, *File System Analysis*, Addison-Wesley Professional, 2005.

[11] B. Carrier and E. Spafford, "Getting Physical with the Investigative Process," *International Journal of Digital Evidence*, vol. 2, no. 2, Fall 2003.

[12] B. Carrier and E. Spafford, "An Event-Based Digital Forensic Investigation Framework," *Digital Forensics Reseach Workshop*, Baltimore, MD., 2004.

[13] P. Carter, “User profiling using text classification,” *Information Assurance Workshop, 2005. IAW ’05. Proceedings from the Sixth Annual IEEE SMC*, june 2005, pp. 274 – 279.

[14] H. Carvey, “The Windows registry as a forensic resource,” *Digital Investigation*, vol. 2, no. 3, 2005.

[15] H. Carvey, “RegRipper,” 2012.

[16] E. Casey, M. Ferraro, and L. Nguyen, “Investigation Delayed is Justice Denied: Proposals for Expediting Forensic Examinations of Digital Evidence,” *Journal of Forensic Sciences*, vol. 54, no. 6, 2009.

[17] E. Casey and G. J. Stellatos, “The impact of full disk encryption on digital forensics,” *SIGOPS Oper. Syst. Rev.*, vol. 42, April 2008, pp. 93–98.

[18] C. Chaski, “Who’s at the Keyboard? Authorship Attribution in Digital Evidence Investigations,” *International Journal of Digital Evidence*, vol. 4, no. 1, 2004.

[19] A. Choudhury, M. Rogers, and W. Gillam, “A Novel Skin Tone Detection Algorithm for Contraband Image Analysis,” *Systematic Approaches to Digital Forensic Engineering*, May 2008.

[20] S. Ciardhuain, “An Extended Model of Cybercrime Investigations,” *International Journal of Digital Evidence*, vol. 3, no. 1, 2004.

[21] F. Dancer and D. Dampier, “A Platform Independent Process Model for Smartphones Based on Invariants,” *Systematic Approaches to Digital Forensic Engineering (SADFE), 2010 Fifth IEEE International Workshop on*, may 2010, pp. 56 –60.

[22] B. Dolan-Gavitt, “Forensic analysis of the Windows registry in memory,” *Digital Investigation*, vol. 5, no. Supplement, 2008, pp. S26 – S32.

[23] D. Farmer and W. Venema, *Forensic Discovery*, Addison-Wesley Professional, 2004.

[24] S. Garfinkel, “Forensic feature extraction and cross-drive analysis,” *6th Annual Digital Forensic Research Workshop*, September 2006.

[25] S. Garfinkel, P. Farrell, V. Roussev, and G. Dinolt, “Bringing science to digital forensics with standardized forensic corpora,” *Digital Investigation*, vol. 6, no. Supplement, 2009, pp. S2 – S11.

[26] W. Gillam and M. Rogers, “File hound: A forensic tool for first responders,” *Digital Forensics Reseach Workshop*, New Orleans, LA, 2005.

[27] A. Grillo, A. Lentini, G. Me, and M. Ottoni, “Fast User Classifying to Establish Forensic Analysis Priorities,” *IT Security Incident Management and IT Forensics, 2009. IMF ’09. Fifth International Conference on*, sept. 2009, pp. 69 –77.

[28] C. Hargreaves and H. Chivers, "Recovery of Encryption Keys from Memory Using a Linear Scan," *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*, 2008, pp. 1369 –1376.

[29] T. Hill and P. Lewicki, *Statistics: Methods and Applications*, StarSoft, Inc., 2005.

[30] R. Ieong, "FORZA – Digital forensics investigation framework that incorporate legal issues," *Digital Forensics Reseach Workshop*, Lafayette, Indiana, 2006.

[31] K. Jones and R. Blani, "Web browser forensics," http://www.securityfocus.com/infocus/1827, 2005.

[32] K. Jones and R. Blani, "Web browser forensics," http://securityfocus.com/infocus/1827, 2005.

[33] R. M. K. Rogers, J. Goldman and T. Wedge, "Computer Forensic Field Triage Process Model," *Journal of Digital Forensics, Security and Law*, vol. 1, no. 2, 2006.

[34] E. E. Kenneally and C. L. Brown, "Revisiting Risk Sensitive Digital Evidence Collection," *Digital Forensics Reseach Workshop*, New Orleans, LA, 2005.

[35] E. E. Kenneally and C. L. Brown, "Risk sensitive digital evidence collection," *Digital Investigation*, vol. 2, no. 2, 2005, pp. 101 – 119.

[36] L. Kenneth, "Child Molesters: A Behavioral Analysis. For Law-Enforcement Officers Investigating Cases of Child Sexual Exploitation Second Edition," Quantico, VA, 1987.

[37] C. King and T. Vidas, "Empirical analysis of solid state disk data retention when used with contemporary operating systems," *Digital Investigation*, vol. 8, no. supplimental, 2011, pp. 111–117.

[38] J. Kornblum, "Preservation of Fragile Digital Evidence by First Responders," *Digital Forensics Reseach Workshop*, Baltimore, MD., 2004.

[39] W. Kruse and J. Heiser, *Computer Forensics : Incident Response Essentials*, Addison-Wesley Professional, 2001.

[40] R. Layton, P. Watters, and R. Dazeley, "Authorship Attribution for Twitter in 140 Characters or Less," *Cybercrime and Trustworthy Computing Workshop (CTC), 2010 Second*, 2010, pp. 1 –8.

[41] C. C. M. Reith and G. Gunsch, "An Examination of Digital Forensic Models," *International Journal of Digital Evidence*, vol. 1, no. 3, Fall 2002.

[42] R. P. Mislan, E. Casey, and G. C. Kessler, "The growing need for on-scene triage of mobile devices," *Digital Investigation*, vol. 6, no. 3-4, 2010, pp. 112 – 124.

[43] K. Nance, B. Hay, and M. Bishop, "Digital Forensics: Defining a Research Agenda," *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, 2009, pp. 1 –6.

[44] J. Oh, S. Lee, and S. Lee, "Advanced evidence collection and analysis of web browser activity," *Digital Investigation*, vol. 8, no. Supplement, 2011, pp. 62 – 70.

[45] A. Orebaugh and J. Allnutt, "Classification of Instant Messaging Communications for Forensics Analysis," *The International Journal of Forensic Computer Science*, vol. 1, 2009, pp. 22–28.

[46] G. Palmer, *A Road Map for Digital Forensic Research*, Tech. Rep., Digital Forensics Research Workshop, Utica, NY, 2001.

[47] J. Riley, D. Dampier, and R. Vaughn, *Time Analysis of Hard Drive Imaging Tools*, vol. 285 of *IFIP International Federation for Information Processing*, Springer, 2008, pp. 335–344.

[48] G. Ruibin, G. Ruibin, C. K. Yun, and M. Gaertner, "Abstract Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework," *International Journal of Digital Evidence*, 2005.

[49] R. Thomons and T. Murachver, "Predicting gender from electronic discourse," *British Journal of Social Psychology*, vol. 40, 2001, pp. 193–208.

[50] C. Yeschke, *The art of investigative interviewing*, 2nd edition, Butterworth-Heinemann, Boston, 2003.

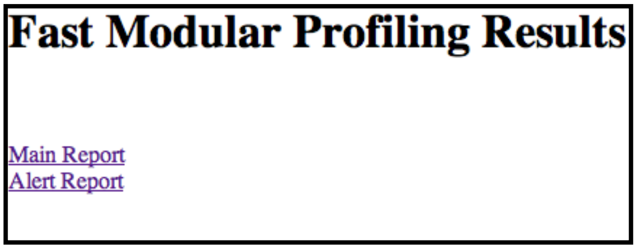
APPENDIX A

SAMPLE FAST MODULAR PROFILING REPORT

The following pages contain a selection from an example FMPU report. The actual report would be in HTML. These pages are only meant to provide an idea of the navigation, and content of a report.

A.1 Index Page

Figure A.1 shows a screen shot of the index page as produced by the FMPU. The index page is automatically opened at the end of the FMPU run. This page links the main report and the alert report.



Fast Modular Profiling Results

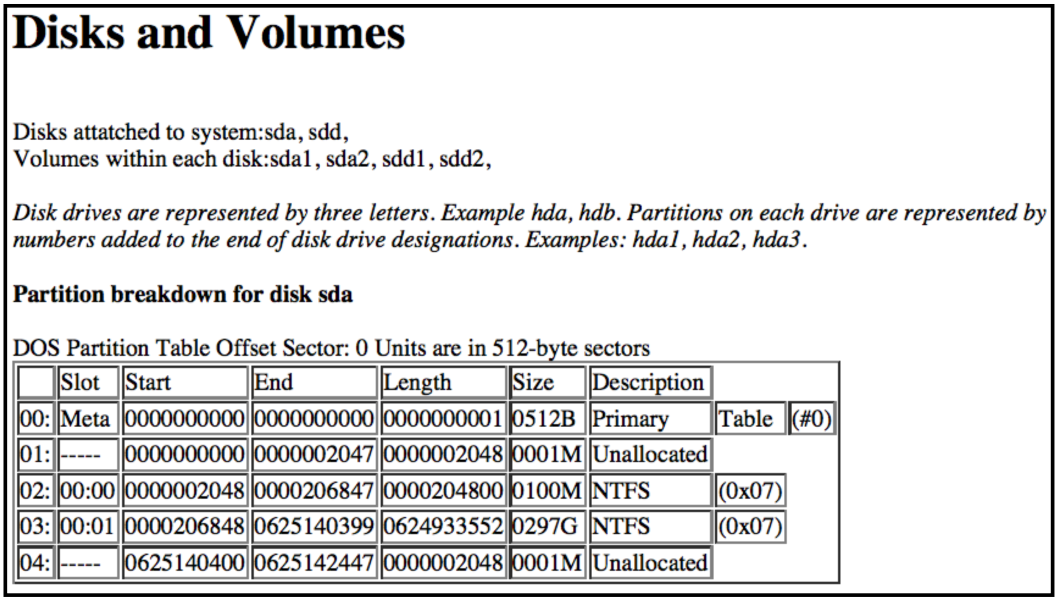
Main Report
Alert Report

Figure A.1

Screen shot of FMPU index page

A.2 Disks and Volumes Report

Figure A.2 displays the Disks and Volumes Report. This report is the first report on the main page. Here the digital triage examiner can quickly identify all physical and logical volumes and view the sector layout of the disk itself.



Disks and Volumes

Disks attatched to system:sda, sdd,
Volumes within each disk:sda1, sda2, sdd1, sdd2,

Disk drives are represented by three letters. Example hda, hdb. Partitions on each drive are represented by numbers added to the end of disk drive designations. Examples: hda1, hda2, hda3.

Partition breakdown for disk sda

DOS Partition Table Offset Sector: 0 Units are in 512-byte sectors

	Slot	Start	End	Length	Size	Description		
00:	Meta	0000000000	0000000000	0000000001	0512B	Primary	Table	(#0)
01:	-----	0000000000	0000002047	0000002048	0001M	Unallocated		
02:	00:00	0000002048	0000206847	0000204800	0100M	NTFS	(0x07)	
03:	00:01	0000206848	0625140399	0624933552	0297G	NTFS	(0x07)	
04:	-----	0625140400	0625142447	0000002048	0001M	Unallocated		

Figure A.2

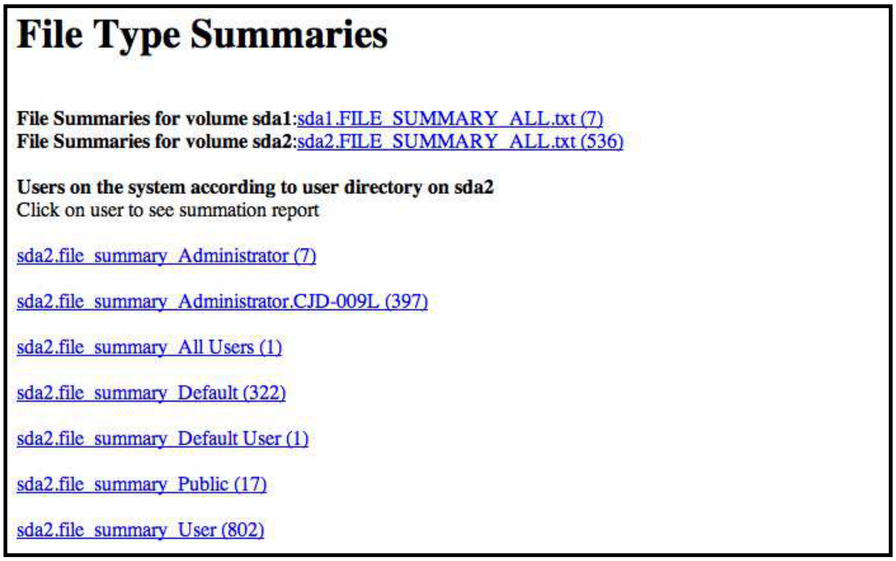
Screen shot of the physical and logical disk layout report

A.3 File Summary Report

The File Summary Report is shown in Figure A.3. This report allows the user to view a listing, by volume or by user directory, of all files on the volume. The number in parenthesis after each link identifies the number of different file types identified. This allows the examiner to quickly determine which volume and which user directory is most used. Figure A.4 shows a sample of the actual file classification report for the entire volume.

A.4 Registry Report

Figure A.5 shows a screen shot of the Registry Report found after the File Summary Report. The first three links shown in Figure A.5 link directly to registry reports for the entire system. The remaining links link to HTML pages outlining the registry reports



File Type Summaries

File Summaries for volume sda1:sda1.FILE_SUMMARY_ALL.txt (7)
File Summaries for volume sda2:sda2.FILE_SUMMARY_ALL.txt (536)

Users on the system according to user directory on sda2
Click on user to see summation report

sda2.file_summary_Administrator (7)

sda2.file_summary_Administrator.CJD-009L (397)

sda2.file_summary_All Users (1)

sda2.file_summary_Default (322)

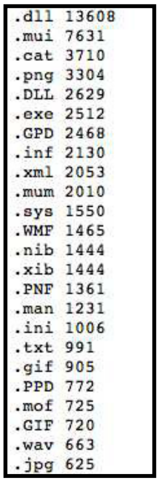
sda2.file_summary_Default User (1)

sda2.file_summary_Public (17)

sda2.file_summary_User (802)

Figure A.3

Screen shot of the file type summaries report



```
.dll 13608
.mui 7631
.cat 3710
.png 3304
.DLL 2629
.exe 2512
.GPD 2468
.inf 2130
.xml 2053
.mum 2010
.sys 1550
.WMF 1465
.nib 1444
.xib 1444
.PNF 1361
.man 1231
.ini 1006
.txt 991
.gif 905
.PPD 772
.mof 725
.GIF 720
.wav 663
.jpg 625
```

Figure A.4

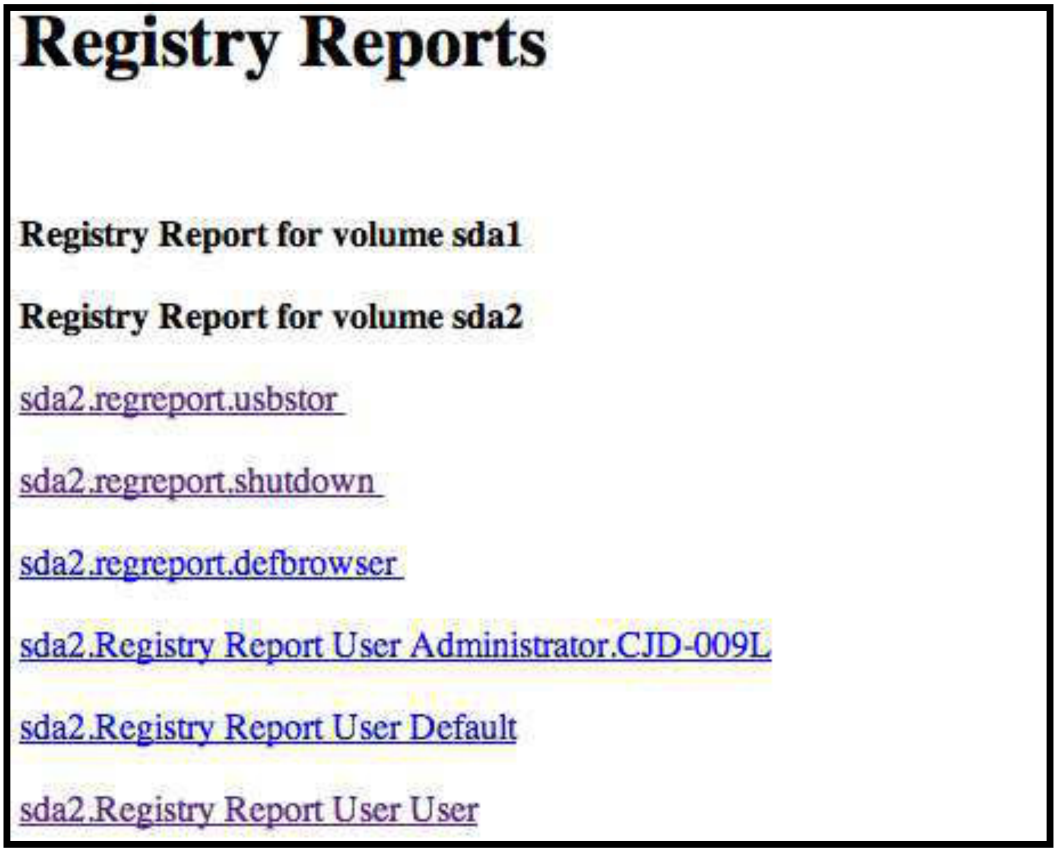
Screen shot of the file type summaries report for a volume generated per user account. Figure A.6 shows the single registry report USBStor. This report shows all USB devices that have been attached to the device recently according to the Windows registry.

A.5 Web History Report

Figure A.7 provides a sample of a Web History Report. The numbers in parenthesis at the end of each line display the number of items contained within the linked report. This allows the digital triage examiner to more quickly sort through the reports to those that have the most items of interest. Figure A.8 shows a sample of a domain summary report.

A.6 Red Flag Report

Figure A.9 shows a Red Flag Report linked from the main page shown in Figure A.1. This report links together all of the results of the keyword filter. Figure A.10 shows a small sample of two file name hits from the keywords: Facebook, and children.



Registry Reports

Registry Report for volume sda1

Registry Report for volume sda2

sda2.regreport.usbstor

sda2.regreport.shutdown

sda2.regreport.defbrowser

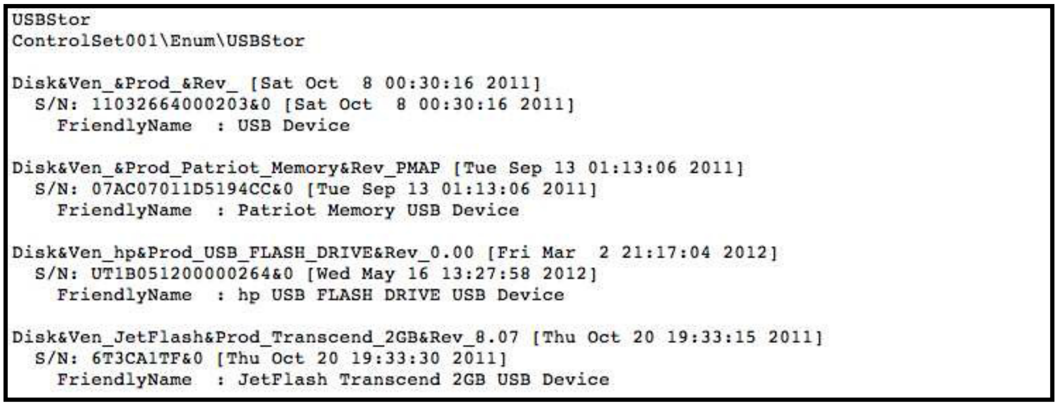
sda2.Registry Report User Administrator.CJD-009L

sda2.Registry Report User Default

sda2.Registry Report User User

Figure A.5

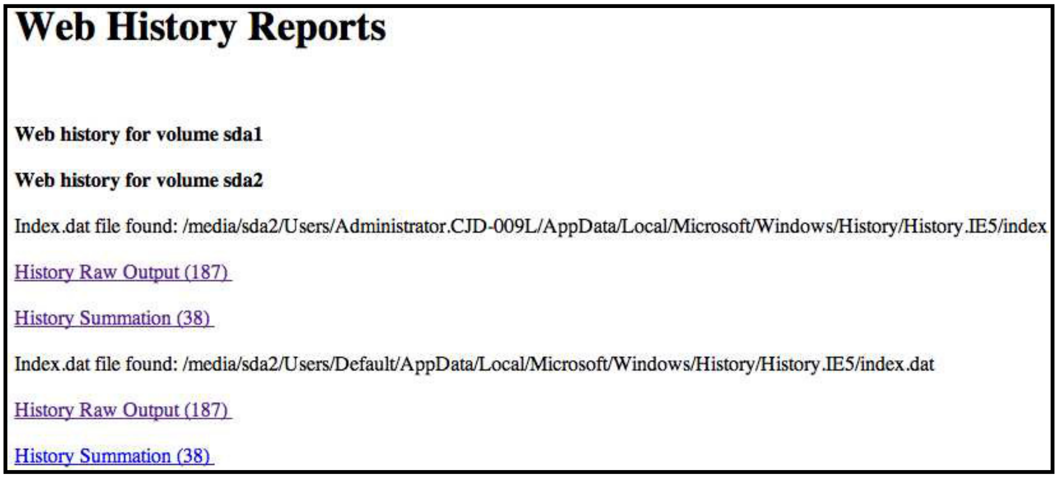
Screen shot of the registry report



```
USBStor
ControlSet001\Enum\USBStor

Disk&Ven_&Prod_&Rev_ [Sat Oct  8 00:30:16 2011]
  S/N: 110326640002036&0 [Sat Oct  8 00:30:16 2011]
    FriendlyName  : USB Device

Disk&Ven_&Prod_Patriot_Memory&Rev_PMAP [Tue Sep 13 01:13:06 2011]
  S/N: 07AC07011D5194CC&0 [Tue Sep 13 01:13:06 2011]
    FriendlyName  : Patriot Memory USB Device

Disk&Ven_hp&Prod_USB_FLASH_DRIVE&Rev_0.00 [Fri Mar  2 21:17:04 2012]
  S/N: UT1B0512000002646&0 [Wed May 16 13:27:58 2012]
    FriendlyName  : hp USB FLASH DRIVE USB Device

Disk&Ven_JetFlash&Prod_Transcend_2GB&Rev_8.07 [Thu Oct 20 19:33:15 2011]
  S/N: 6T3CA1TF&0 [Thu Oct 20 19:33:30 2011]
    FriendlyName  : JetFlash Transcend 2GB USB Device
```

Figure A.6

Screen shot of the USBStor registry report



Web History Reports

Web history for volume sda1

Web history for volume sda2

Index.dat file found: /media/sda2/Users/Administrator.CJD-009L/AppData/Local/Microsoft/Windows/History/History.IE5/index

History Raw Output (187)

History Summation (38)

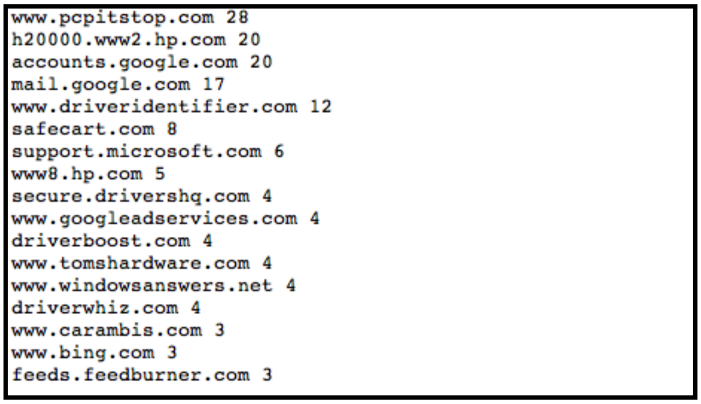
Index.dat file found: /media/sda2/Users/Default/AppData/Local/Microsoft/Windows/History/History.IE5/index.dat

History Raw Output (187)

History Summation (38)

Figure A.7

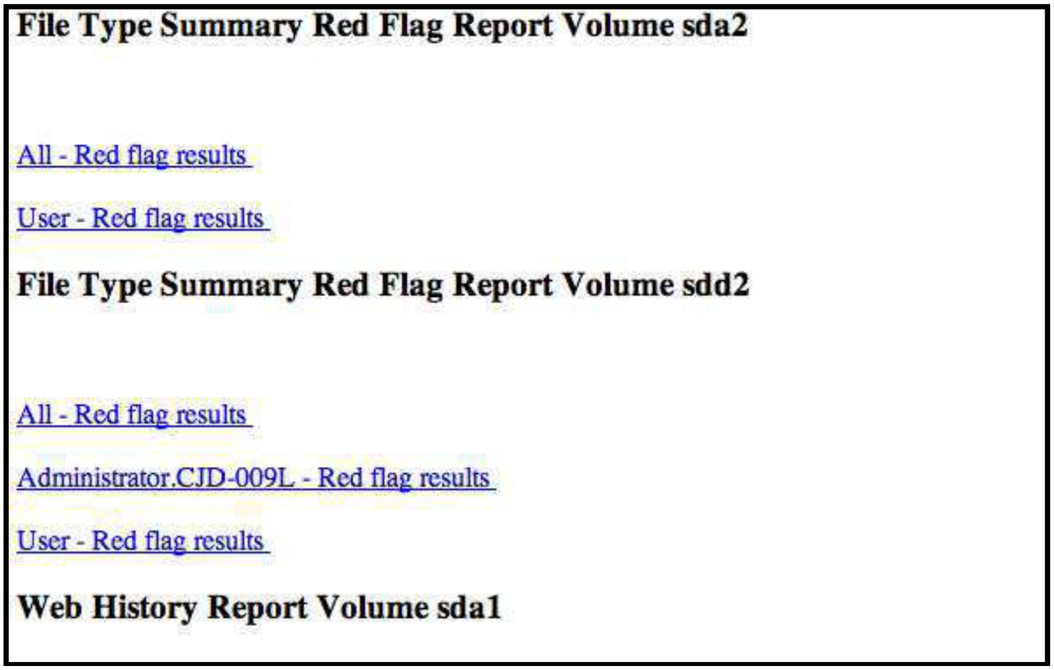
Screen shot of the Web history report



```
www.pcpitstop.com 28
h20000.www2.hp.com 20
accounts.google.com 20
mail.google.com 17
www.driveridentifier.com 12
safecart.com 8
support.microsoft.com 6
www8.hp.com 5
secure.drivershq.com 4
www.googleadservices.com 4
driverboost.com 4
www.tomshardware.com 4
www.windowsanswers.net 4
driverwhiz.com 4
www.carambis.com 3
www.bing.com 3
feeds.feedburner.com 3
```

Figure A.8

Screen shot of the Web history report domain summary



File Type Summary Red Flag Report Volume sda2

All - Red flag results

User - Red flag results

File Type Summary Red Flag Report Volume sdd2

All - Red flag results

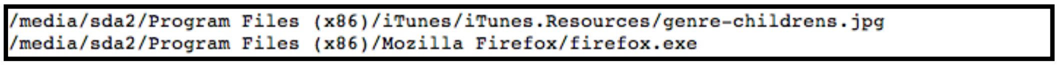
Administrator.CJD-009L - Red flag results

User - Red flag results

Web History Report Volume sda1

Figure A.9

Screen shot of the red flag report



```
/media/sda2/Program Files (x86)/iTunes/iTunes.Resources/genre-childrens.jpg
/media/sda2/Program Files (x86)/Mozilla Firefox/firefox.exe
```

Figure A.10

Screen shot of the red flag report with two file name hits