Simplified and Inexpensive Mobile Digital Forensic Device

A Thesis Presented to

The Faculty of the Computer Science Department

California State University Channel Islands

In (Partial) Fulfillment

of the Requirements for the Degree

Masters of Science in Computer Science

by

Eric Elwood Gentry

Advisor: Michael Soltys

December 2018

APPROVED FOR MS IN COMPUTER SCIENCE

Advisor: Advisor Name **Date**

Name **Date**

Name **Date**

APPROVED FOR THE UNIVERITY

Name **Date**

Non-Exclusive Distribution License

In order for California State University Channel Islands (CSUCI) to reproduce, translate and distribute your submission worldwide through the CSUCI Institutional Repository, your agreement to the following terms is necessary. The author(s) retain any copyright currently on the item as well as the ability to submit the item to publishers or other repositories.

By signing and submitting this license, you (the author(s) or copyright owner) grants to CSUCI the nonexclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that CSUCI may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that CSUCI may keep more than one copy of this submission for purposes of security, backup and preservation.

You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. You also represent and warrant that the submission contains no libelous or other unlawful matter and makes no improper invasion of the privacy of any other person.

If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant CSUCI the rights required by this license, and that such third party owned material is clearly identified and acknowledged within the text or content of the submission. You take full responsibility to obtain permission to use any material that is not your own. This permission must be granted to you before you sign this form.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN CSUCI, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

The CSUCI Institutional Repository will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Title of Item

---

3 to 5 keywords or phrases to describe the item

Author(s) Name (Print)

Author(s) Signature Date

Simplified and Inexpensive Mobile Digital Forensic Device

Eric Elwood Gentry

May 24, 2018

Abstract

Contents

List of Figures

1 Abstract

Keywords: Digital Forensics, Triage, Mobile, Digital Evidence Backlog, Raspberry Pi

Goals:

- preserve and protect evidenciary integrity
- reduce evidence gathering and triage analysis time
- prevent adding more to backlog than necessary by preventing over-confiscation
- reduce need for on-scene Digital Forensic Ssientists
- reduce backlog of digital evidence for tackling backlog

SEAKER tradeoffs: Precision (only relevant files) vs Recall (all relevant files) - level of recall required at triage stage can be sacraficed

Introduce online storage system for digital forensic metadata format to enhance sharing capabilities across jurisdiction boundaries and prevent sharing complexities

2 Introduction

I know you are wondering what I am going to say here. Your guess is as good as mine. I really like the introduction section, because I can say whatever I want! LOL

3 Meetings with Frank

3.1 May 18th, 2018

My meeting with Frank Lyu, a civilian working for the Ventura County Sheriff's department, on loan to the Southern California High Tech Task Force

(SCHTTF) went well. The SCHTTF is an 8 person team made up of four civilians and four deputies, all reporting directly to the Ventura County District Attorney's office. We met for lunch and I was able to ask him questions about the environment he works in as well as touch on ranking the high value items that this SEAKER project could provide.

First, we talked about his work environment. He has several responsibilities working for SCHTTF. The first and foremost is his caseload, which consists of examining digital evidence using forensic techniques in his lab that result in a report to the District Attorney's office. His other responsibilites include assisting the District Attorney and staff with evaluating defense evidence reports, studying digital forensic technologies (for when he has to explain things to juries), helping other agencies identify and catalog evidence that they are not familiar with, helping serve warrants on critical cases, and helping to retrieve lost digital materials for other law enforcement agencies.

He explained that following the processes and procedures is by far the most important aspect of his job. The evidence handling, storage, and evaluation are critical to whether a case succeeds or is dismissed. Frank began to describe the intake process, which involves the evidence, the agency report, and the search warrant that will be used to search and evaluation the materials. We did not go into further detail.

Items he looks for during an on-site warrant are: user accounts, previewing the materials (especially in cases involving CP), and checking for the existence of Peer-to-Peer sharing utilities. In addition, he strives to collect the following networking information: publically broadcast SSIDs, each SSID's level of encryption, how many devices are connected to the router, and the external IP address for the router.

Value of SEAKER in the field: Filename search utilizing regular expression, the networking information specified in the last paragraph, producing a report, making an ios app instead of using a webpage, links for the files found, and clickable thumbnails.

Value of SEAKER in the lab: Filename search utilizing regular expression and SEAKER hashes of images. He also mentioned a Microsoft tool called PhotoDNA that they currently use to find naked humans.

I asked about obtaining all of the statistics related to ingestion or evidence, caseload, pace at which evidence can be evaluated, etc. Frank's answer was that Adam could probably provide that information without lab access, but that Michael should be asked to talk to Adam.

Finally, Frank mentioned that sometimes in financial cases, he is asked to search computers at business offices, and the ability to search for specific filenames is very important there.

4 Background

4.1 Review Material and Analysis

Current challenges and future research areas for digital forensic investigation - Lillis et al., 2016 [7]

Some of the current challenges in digital forensic investigations are directly related to the amount of data being created. As Lillis et al[7] explores in their research, there are three main factors involved in the digital forensic backlog: increasing number of devices seized per case, increased number of cases involving digitial evidence, and the increasing volume of data per digital media. This has lead to a growing and already substantial backlog in digital forensic investigations.

One effect of this increased delay and backlog is that cases become inactive, waiting for new leads. A more aggressive approach to solving the backlog could help prevent dismissals, cold cases, and potentially more societal harm from a corrupt investigation suspect.

Raghavan[10] has accumulated a list of 5 major challenges that the digital forensics community is facing and continue to add to the backlog problem.

The first is the complexity of binary data aquisition, i.e. low level data aquisition through digital media duplication. This challenge causes the need for sophisticated data reduction techniques.

Another complexity is the diversity of data and lack of standard examination techniques. The plethora of operating systems and file formats has been increasing and is posing a more and more significant challenge over time.

The consistency and correlation problem is yet another challenge. This is a problem resulting from the current digital media investigation tools not providing the entire picture to investigators. Only part of the whole picture is provided when these tools find digital evidence.

Another issue that Raghavan[10] proposed is the volume of data to sort through. The sheer amount of data that exists per user is increasing at an alarming rate [cite?], and has lead to a very large backlog of digital evidence to investigate. These delays have even caused some cases to be dismissed. This challenge is exacerbated by the lack of adequate automation for digesting the data.

The fifth, but certainly not the last, challenge proposed by Raghavan[10] is the timeline synchronization issue with digital evidence. Since the evidence could be collected in different time zones, with different timestamp formats, clock skew, etc, lining up the events in order can be challenging or infeasible.

With the proliferation of Internet Of Things (IOT) devices and cloud storage, the field of digital forensics continues to expand. These areas pose a great challenge, but also new opportunities. Lillis et al[7] researched cloud storage and found some areas of opportunity, for instance parallel processing, distibuted computing, GPU/FPGA utilization, and others. These areas for increasing the efficiency of ditigal forensics can be explored further due to the substantially reduced I/O limitations in cloud storage.

The Internet of Things (IOT) also poses new challenges. IOT devices are estimated to number near 40 billion by 2020, contributing to the overwhelming amount of digital data. Since these devices tend to have more non-persistent memory and less storage, this causes added complexity for gathering and analysis. In addition, a portion of IOT devices are battery operated and computationally challenged, leading to loss of data over time.

Tiered forensic methodology model for Digital Field Triage by non-digital evidence specialists - Hitchcock et al., 2016[4]

Hitchcock et al[4] has proposed and evaluated a "tiered forensic methodology" model that defines a process of digital forensic triage utilizing non-digital evidence specialists. In their research, they identified a large and growing backlog of digital evidence. This backlog has led to problems in the law enforcement community with regards to collecting, analyzing, reporting, and prosecuting.

The summary of the research done by Hitchcock et al[4] are as follows. They sought to expedite the process of sending digital evidence for analysis and results. One of their goals is to enable more field triage of digital evidence to reduce the amount collected, and act specifically on pertinent information only. They recommended that some front-line crime scene investigators (non-forensic analysts) be trained in the implementation of digital evidence triage and evaluation. These trained individuals would be Digital Field Triage (DTF) experts and have the ability perform field-level digital evidence triage. This triage would specifically weed out the benign from the consequential digital evidence with high certainty, while also protecting the digital evidence from spoilage and preserving evidentiary integrity.

The next tier is where the already-triaged digital evidence is sent for full evaluation. This is a certified facility that can perform full digital forensic analysis, called a Technological Crime Unit (TCU). The TCU is currently heavily inundated with cases needing analysis and reporting of digital evidence.

The backlog and delays in case reporting are contributors to a common problem of time sensitivity. Some countries have given their citizens a right to a "speedy" trial. As well, some countries have statutes of limitation (limits on how long after the crime was committed to resolve the case) for most crimes. Some administrative situations are also contributors, for instance case prioritization based on chronological filing, crime severity, or victim needs.

This tiered approach is based on a Computer Forensic Field Traige Process Model proposed by Rogers et al [11] and the internation standard ISO 27037 (Information Technology - Security Techniques - Guidelines for identification, collection, aquisition, and presentation of digital evidence). The process model breaks down the six phases of digital evidence categorization, which Hitchcock et al[4] loosely based their four phase approach on. The four phases are: planning, assessment, reporting, and threshold. The ISO 27037 standard specifically attempts to address the need to minimize the risk of potential digital evidence being spoiled by mis-handling, while also attempting to maximize the evidentiary value of digital evidence collection.

This approach is not without risks. One concern is the accidental exclusion of an item of digital evidence that is important to the investigation. Another is the level of computer skills and training of the DTF expert. The paper does attempt to mitigate the latter with training and management process, while providing evidence that the former is a common misconception in most cases.

The research of Hitchcock et al[4] should be referenced for a good process starting point for digital forensic labs.

A practical and robust approach to coping with large volumes of data submitted for digital forensic examination [13]

One digital evidence triage method proposed by Shaw et al[13] seeks to standardize on an approach they call "enhanced previewing". Enhanced previewing seeks to solve some of the problems associated with typical triage approaches. As is the case in other research, Shaw et al[13] extolls the need to reduce digital forensic evidence analysis backlogs, especially with the evolution of big data and the proliferation of digital devices.

Shaw et al[13] points out that neither digital forensic triage examination nor digital forensic full examination are well defined. Triage may mean something completely different to two digital forensic examiners. As well, full digital forensic examination has no robust standard to follow.

The proposal for a practical and robust method by Shaw et al[13] aims to stem the concerns of a typical triage process. Risks still exist, for instance overlooking digital evidence, but it is argued that those risks are outweighed by the risks of a lengthy process due to large backlogs and the associated delays in evaluating that evidence. Another concern esists that inadequately trained people will be charged with performing on-site digital evidence triage and mishandling or incorrectly evaluating results will cause evidence spoilage. Other concerns are the potential high cost of software and training.

In order to provide a simple, yet robust mechanism, Shaw et al[13] starts with an open source, CD-bootable image of GNU/Linux and enhances its features to include boot-time application launching, and a simple to use interface with minimal ability to deviate from task. This bootable CD is intended to be placed into evidenciary computer systems and booted using a series of BIOS modifications or boot-time interruptions. This mechanism to boot the system off of a bootable CD is difficult, and where the most problems with untrained users of the enhanced previewing will happen.

The enhanced previewing concept has valuable merit, in that the collection mechanisms are thorough. Using the GNU/Linux based system and having written code for it, Shaw et al[13] utilized some well thought-out approaches. First, all hard drives from the evidentiary system are mounted into the GNU/Linux filesystem as read-only, thereby obviating the need for write-blockers. As well, the entire hard drive is evaluated, including the file system, all partitions, unallocated space, deleted files, and compressed files. In addition, other mechanisms are employed that continue to enhance the previewing are employed. One example is to note encrypted compressed files for review later.

They also strongly compliance with digital forensics best practices, like the ones provided in the ACPO. This standards guide is from the UK and is titled Best Practices Guide for Digital Evidence (2012). There are four main practices outlined, but the main one that Shaw et al[13] chose to point out was the second one regarding the integrity protection of the original digital media.

The design science research process: a model for producing and presenting information systems research - Peffers et al., 2006[9]

Testing the harmonised digital forensic investigation process model-using an Android mobile phone [8]

Forensic analysis of iPhone backups [12]

Forensic analysis of social networking applications on mobile devices [1]

Jailbroken iPhone Forensics for the Investigations and Controversy to Digital Evidence [2]

Methods and tools of digital triage in forensic context: survey and future directions [6]

A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview [5]

Forensic examination of digital evidence: a guide for law enforcement [3]

5 Analysis

5.1 Process for gathering digital evidence

First of all, never boot a computer. That will alter the original state of the hard drive due to the operating system being loaded. This may alter the data available for collection and render the digital media "spoiled" and therefore unusable ad evidence in a case.

The main point here is to preserve evidentiary integrity and protect against spoilage.

Specific training is required to be able to handle digital evidence properly. What do digital evidence collectors already have to go through in terms of training?

5.2 SEAKER Usage Methodologies

5.2.1 Connected Method

In this approach, SEAKER is hard-wired using an ethernet cable to the internet or the lab intranet. The connection is utilized to connect directly to the Image Hash Storage Server (IHSS) and the Digital Evidence Storage Server (DESS). The digital evidence is still collected locally, but also being transmitted to the DESS and comparing image hashes to the IHSS.

5.2.2 Disconnected Method

In this approach, SEAKER is not connected to an ethernet cable and is solely being used as a wifi router for collecting digital evidence locally. A unique digital collection ID will be created when connected to the DESS at a later time.

5.3 Automated processes during SEAKER evaluation

5.3.1 Local processing

This section describes the local processing that takes place in both Connected and Disconnected Methods.

- picture of digital media
- picture of hosting hardward platform (laptop, computer, server, phone, etc)
- file list
- full log of capturing/viewing/analysis
- image thumbnails

- video thumbnails
- browser history
- emails
- user profiles
- deleted files
- image thumbnail subsets (images with faces, bodies, documents, etc)
- searches performed
- anything ”marked” as an ”artifact”

5.3.2 Remote processing

This section describes the remote processing that takes place only when Connected Method is in use.

- matching hashes of images
- matching hashes of videos
- online storage of digital evidence collection for this case at this site and time

6 Conclusion and future work

My conclusion is going to be very fascinating and wonderful. I can feel it.

6.1 Future work

- Online hard drive investigation (i.e. Cloud Forensics)
- Network Traffic Investigation
- Video segmentation and video image hashing
- crime-specific searchs:
  - financial crimes
  - credit card fraud
  - hacking
  - bullying
  - bloackmail
  - espianage
  - fraud
  - customizable (corporate / military)
- OS lockdown (raspbian)
- phone specific OS tools
- phone specific apps
- encrypted devices (password entry location, assessment without password)
- running machine RAM assessment
- utilize forensics as a service

7 Code improvements

- Query Expansion - automatically searching for same query maybe other contexts
- Synonym Matching - automatically searching for similar words to query word
- collect everything in UTC time
- universal way of collecting hard drive hash for verification of evidence integrity
- Data Visualizations:
  - present all data visualizations for particular drive or all hard drives
  - graph - size vs amount of files (one hard drive, and all hard drives)
  - graph - common details (like file type, etc) maybe clickable!
  - graph/chart - files by date
  - graph/chart - files by file type
  - chart - website visits
  - digital image hashes list (stored and compared)
- for analysis: skip OS files, applications files, etc
- lock down OS
- Investigation Gathering rollup: (stored online)
  - Database Schema
  - metadata
  - Unique "gathering ID"
  - case number
  - observation report
  - crime severity
  - potential offenses
  - time gathered

– gatherer
– suspect list
– location gathered
– suggestions for other research
– which computer system it came from
– SET of evidence
– Digital Evidence item
– images of item
– unique item ID
– file contents
– ranking within set of evidence
– image thumbnails
– collection statistics
– etc

8 Keywords and glossary

- Contraband files
- stages: preprocessing, storage, analysis, reporting
- stages: gather (document, catalog), triage (analyze, live review, automated review, artifact storage, meet threshold?), results (present, graphs, search)
- Chain of Custody for digital evidence
- Forensic integrity of digital evidence
- artifacts = pieces of digital evidence that are of importance to the case
-

9 Graphs, Images, Figures, and tables

- figure - field triage flowchart
- figure - each stage field triage flowchart
- figure - Image analysis and hash creation flowchart
- graph - aquisition time vs full
- graph - investigation time vs full
- graph - analysis time vs full
- graph - total time from initial plug-in to decision to be evidence (threshold)
- graph - current backlock in ventura county
- figure - DFT vs TCU (Hitchcock[4])
- graph - collection time vs number and size of files

References

[1] Noora Al Mutawa, Ibrahim Baggili, and Andrew Marrington. Forensic analysis of social networking applications on mobile devices. *Digital Investigation*, 9:S24–S33, 2012.

[2] Ya-Ting Chang, Ke-Chun Teng, Yu-Cheng Tso, and Shiuh-Jeng Wang. Jailbroken iphone forensics for the investigations and controversy to digital evidence. *Journal of Computers*, 26(2), 2015.

[3] Sara V Hart, John Ashcroft, and Deborah J Daniels. Forensic examination of digital evidence: a guide for law enforcement. *National Institute of Justice NIJ-US, Washington DC, USA, Tech. Rep. NCJ*, 199408, 2004.

[4] Ben Hitchcock, Nhien-An Le-Khac, and Mark Scanlon. Tiered forensic methodology model for digital field triage by non-digital evidence specialists. *Digital Investigation*, 16:S75–S85, 2016.

[5] Joshua I James and Pavel Gladyshev. A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. *Digital Investigation*, 10(2):148–157, 2013.

[6] Vacius Jusas, Darius Birvinskas, and Elvar Gahramanov. Methods and tools of digital triage in forensic context: survey and future directions. *Symmetry*, 9(4):49, 2017.

[7] David Lillis, Brett Becker, Tadhg O'Sullivan, and Mark Scanlon. Current challenges and future research areas for digital forensic investigation. *arXiv preprint arXiv:1604.03850*, 2016.

[8] Stacey Omeleze and Hein S Venter. Testing the harmonised digital forensic investigation process model-using an android mobile phone. In *Information Security for South Africa, 2013*, pages 1–8. IEEE, 2013.

[9] Ken Peffers, Tuure Tuunanen, Charles E Gengler, Matti Rossi, Wendy Hui, Ville Virtanen, and Johanna Bragge. The design science research process: a model for producing and presenting information systems research. In *Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006)*, pages 83–106. sn, 2006.

[10] Sriram Raghavan. Digital forensic research: current state of the art. *CSI Transactions on ICT*, 1(1):91–114, 2013.

[11] Marcus K Rogers, James Goldman, Rick Mislan, Timothy Wedge, and Steve Debrota. Computer forensics field triage process model. In *Proceedings of the conference on Digital Forensics, Security and Law*, page 27. Association of Digital Forensics, Security and Law, 2006.

[12] B Satish. Forensic analysis of iphone backups. *Securitylearn. net.*

[13] Adrian Shaw and Alan Browne. A practical and robust approach to coping with large volumes of data submitted for digital forensic examination. *Digital Investigation*, 10(2):116–128, 2013.