SEAKER:
A Mobile Digital Forensic Triage 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

SEAKER:
A Mobile Digital Forensic Triage Device

Eric Elwood Gentry

July 4, 2018

Keywords: Digital Forensics, Digital Forensics Triage, Mobile Digital Forensics, Digital Evidence, Digital Evidence, Forensic Tools, Raspberry Pi

Abstract

As our world of digital devices continues to expand, the potential for digital evidence available to law enforcement during case investigation is ever increasing. The growing amount of digital evidence, along with the deprived pool of Digital Forensic Investigators is causing a backlog to form at many of the digital forensics labs around the world. This backlog leads to delays in evidence analysis and reporting, causing investigators and prosecutors to postpone or even drop on-going cases.

The SEAKER device is a digital forensic triage tool that is designed to be simple, portable, inexpensive, robust, and easy to use. SEAKER is an acronym for Storage Evaluator And Knowledge Extraction Reader. Utilizing a raspberry pi, this is a novel approach to helping provide immediate feedback to investigators along with attempting to stem the backlog problem. It was originally developed for on-scene investigations that require immediate feedback, especially in time-sensitive investigations. It also appears to be an excellent tool to help reduce the backlog by preventing over-collection of digital evidence. SEAKER is not meant to replace a fully-functional digital forensic lab, but instead to augment the initial investigation and help reduce the backlog. This research and device overview proposes the mobile, inexpensive, digital triage device called SEAKER.

Contents

List of Figures

List of Tables

1 Introduction and Literature Review

1.1 Introduction

Law enforcement investigations involve many aspects of criminality and need carefully thought-out procedures and practices. These procedures and practices are essential to finding the evidentiary information necessary to determine criminal liability, but are also in place to ensure that the evidence collected is not tainted and is sound, viable, admissible court evidence. Establishing and retaining the forensic integrity of the evidence is a required and crucial part of the investigator's task.

Performing investigations is also a noteworthy endeavor. There are many steps involved that require special training to be performed properly. One primary example is the *chain of custody*. This refers to the step-by-step documentation record regarding evidence that includes details such as who had custody of the evidence, when they had custody, who it was transferred to, who analyzed it, etc. Another is the exacting science of collecting, labeling, itemizing, and acquiring of evidence. For instance, collecting physical evidence requires the use of gloves, evidence bags, fingerprint-dusting equipment, etc. to prevent cross-contamination, fingerprint smudging, DNA evidence mishandling, and a multitude of other evidence tainting. Without the proper adherance to guidelines, even conclusive evidence may not be admissible during a case.

Digital evidence is also very essential to many investigations and cases in the modern world. With each passing year, more and more digital devices are collecting, storing, and uploading data. As well, electronic devices for personal use appropriately labelled the Internet of Things (IoT) or the Internet of Everything (IoE) are becoming more and more ubiquitous in our everyday lives. IOT devices are now everyday household items like refrigerators, thermostats, light bulbs, window coverings, garage door openers, keys, clothes, and much more. These devices and the massive amount of digital information that is being generated and collected are often helpful in criminal investigations. The data can be used to construct timeframes of activity, locations of individuals, Internet activity, computer users and usages, and lots other potential digital information.

One growing and particularly helpful aspect of an investigation is digital forensics. This not only involves collecting potential digital evidence, but also analyzing, and reporting procedures. This almost always requires a search warrant - a court-ordered search and seizure of potential evidence of a location where a suspect resides, works, or may be storing it. The search warrant is executed after it has

been obtained from a judge and can involve physical and digital evidence, as well as other items of consequence.

Search warrant investigations are often fraught with danger, intentional obscurity, hidden evidence, and potential mishandling of evidence. Before anything else can be done, the location must be considered secure - considered safe from harming investigators and free of potential threats. Once a scene is secured at a search warrant service involving electronic evidence, three activities take place simultaneously: the search for physical evidence, the search of the physical evidence itself for electronic evidence, and the interviewing of involved parties.

The physical and digital evidence can guide the interviewing of the suspect(s), but also has the potential to have both positive and negative effects on the outcome of the investigation. If investigators do not locate any physical evidence for an examiner to evaluate, then intelligence is not gathered and the interviewer has less information with which to confront the suspect(s). If investigators present physical evidence to an examiner who is able to evaluate it quickly in the field, then the interviewer (who is oftentimes also the lead investigator on the case) can confront the suspects and potentially secure statements that lead to prosecution.

This leads to the need for digitial forensics specialists to bring their lab equipment into the field, especially when serving a search warrant. The lab equipment is specialized software and hardware designed to analyze, report, and maintain forensic integrity on potential digital evidence. This equipment often involved a laptop, a write-blocking device, media imaging storage devices, expensive software, and associated cabling for connection and power. As well, this software is designed for extensive and in-depth searching and often takes hours or days to analyze the evidence. Many of the reports from these systems are designed to be thorough and may take a skilled digital forensic examiner days to pour over the material produced. Oftentimes, this equipment is not brought into the field and all digital evidence is simply collected for later analysis at the law enforcement facilities.

The need for a more field-friendly digital forensic *triage* solution will assist in the initial investigation tasks in multiple ways:

1. It enforces a structured procedure and approach that is user-friendly to *non-digital forensic aware investigators* with the goal of simple instructions for use and very simple evidence location.

2. It enables investigators, especially interrogators, a very fast digital-evidence overview into the types of files and information being accessed and stored

on the computer equipment at the site of the search warrant.

3. It limits the number of devices and therefore the amount of data required in the in-depth analysis phase at the lab.

4. It minimized the impact and inconvenience to innocent parties at the site of the search warrant. The devices that are searched and found to have no evidenciary value can be deemed inconsequencial to the case and remain at the scene.

5. It may be used to provide initial, albeit simplified, analysis results on potential digital evidence ingested into the digital forensic lab.

The topic explored in this paper is a portable, inexpensive, efficient device, named SEAKER, that is intended to overcome the need for a full digital forensic lab equipment suite to be brought into the field. The SEAKER device was conceived and an initial prototype was produced at the California State University Channel Islands (CSUCI) campus in a Masters level Computer Security class (COMP 524, Summer 2017) in direct collaboration with the Southern California High Tech Task Force (SCHTTF) division of the Ventura Country District Attorney's (VCDA) office.

1.1.1 Author's Contributions

Contributions I have made to the SEAKER device project:

- Developed the SEAKER creation bash script to turn a standard Raspberry Pi into a SEAKER by programmatically installing raspbian software packages, setting up WIFI as a wireless access point, adding a web server, and preparing the running environment with the proper fileset

- Wrote custom C program to increase device searching efficiency in lieu of slower, native operating system solutions for finding content on digital media

- Co-presented and demonstrated the initial SEAKER prototype for the Summer 2017 Masters level CSUCI Security class project to SCHTTF, CSUCI department heads, and local community leaders

- Presented SEAKER as a thesis project at the April 2018 CSUCI Cyber Security Event to CSUCI President Erika Beck, California State Assembly Member Jacqui Irwin, and local community leaders

- Co-authored a conference paper on the SEAKER project and the technology behind it
- Updated and enhanced SEAKER functionality to support the latest raspbian operating system (Stretch Lite, April 2018 release), including enabling ethernet passthrough to the wireless access point
- Co-authored by creating Logic Models and assisting with read-throughs for a United States Department of Justice (DOJ) grant proposal for future work on this project (SEAKER) and a second, related security project (Voyager)

1.2 Literature Review

1.2.1 History of Digital Evidence

The digital forensics field began in the mid 1980s with an understanding from several law enforcement agencies that computers would play a critical role in criminal investigations of the future. In 1993, the FBI hosted an international conference on computer evidence in Virginia. This was the first major conference on the subject and had attendees from 26 different countries. Much of the original computer forensics at that time related to recovering information from local computers.

Among the early pioneers in the digital forensics field, there was a common understanding that a system of processes and procedures were needed to locate, record, analyze and report information. This process would have to be similar to how non-digital physical evidence was handled, but also include other computer specific preservation methods to ensure the integrity of evidence found. Those processes and procedures have increased in complexity over time and have suffered from the lack of ubiquitous adoption to a single standard.

Even as late as 2013, Shaw et al[6] points out that neither digital forensic triage examination nor digital forensic full examination are well defined. Digital forensic triage may mean something completely different to two digital forensic examiners. As well, full digital forensic examination has no robust standard to follow, Although there has been no shortage of attempts.

In 2006, Rogers et al[5] proposed a standardization model for a portion of the entire process called *triage* for digital forensic examiners to follow: Computer Forensics Field Triage Process Model. The authors of the CFFTPM note the important legal and technical considerations prior to implementing CFFTPM on a particular investigation. The legal considerations include issues related to search

warrant scope and limitations, U.S. Constitutional 4th Amendment rights, etc. The technical considerations include type of case, criticality of timeliness, skillset of the on-site digital forensic examiner, skillset of the suspect, having proper lab equipment on-site, scene control, etc.

The National Institute of Standards and Technologies (NIST) published guidelines for several different types of digital evidence, for instance mobile phones, and computers. However, NIST focuses on the analysis portion of the science, but leaves the collection, and reporting aspects unexplored. The International Standards Organization also published a set of guidelines, but primarily focused on collection and handling aspects of digital evidence. Ajijola et al[1] provided a thorough review in 2014 of the NIST SP 800-101 Rev. 1:2014 guidelines titled Guidelines on Mobile Devices Forensics and ISO/IEC 27037 titled Guidelines for Identification, Collection, Acquisition, and Preservation of Digital Evidence. Their recommendation was a combined approach, though still not a fully formed solution.

1.2.2 Process Standardization

Several approaches over the years have been proposed as universal processes and procedures to gathering, reviewing, and presenting digital evidence. These approaches range in number of steps, process coverage, and overall method, but all have the common goal of finding usable digital evidence for preventing future harm to society and preserving its integrity for means of presentation in court cases.

The Association of Chief Police Officers (ACPO), a private company that helped establish and develop policing practices in England, Wales, and Northern Ireland for many years, put together a *Good Practice Guide for Digital Evidence*[7] in 2012 that outlines some recommended procedures for dealing with digital evidence. As with other methodologies, this guide explains using a four step approach: Plan, Capture, Analyze, Present.

Other researchers have recommended similar approaches. Shaw et al[6] analyzed the ACPO and focused on the second step (Capture) as the primary guideline for evidential integrity. They strongly suggest compliance with digital forensics best practices, like the ones provided in the ACPO. However, their final recommendation

Rogers et al[5] proposed a reliable, repeatable process model in 2006 for digital evidence triage in the field. It was created in partnership with Purdue University

Cyber Forensics and Computer and Information Technology Departments, along with the National White Collar Crime Center ¡NEED CITE¿. The process is derived from several other models including Integrated Digital Investation Process model (IDIP), Digital Crime Scene Analysis (DCSA), and a military Operations Order (OpOrd). In coordination with the Southern Indiana Assistant U.S. Attorney's office, USADA Steve Debrota, Rogers et al[5] implemented and reported on the success of their proposed model: Computer Forensics Field Triage Process Model (CFFTPM).

Rogers - In addition, the process is intended to protect the integrity of the evidence or potential evidence in an investigation.

1.2.3 Aquisition Methodology

Ajijola - The ISO guidelines provide an international accepted approach, making it easier to compare, combine, and contrast results for out-of-jurisdiction cases and for data scientists research. It provides a common reference line for digital forensics. However, it is not meant to replace laws or regulations. The main purpose is to provide practical assistance for investigations involving potential digital evidence, while preventing digital evidence corruption. This process facilitates the usability of evidence by other jurisdictions. This guideline provided four steps for handling potential digital evidence: Identification, Collection, Acquisition, and Preservation. However, this is incomplete, as it only addresses gathering, not actually evaluating or providing results to law enforcement investigators.

1.2.4 Analysis Methodology

Rogers- The primary machine type that this model covers is the standard Windows machine.

Ajijola et al[1] - The NIST guidelines provide an in-depth look into mobile devices, helping to explain the technology involved and its relationship to the forensic process. NIST itself is a technological, non-regulatory federal agency under the U.S. Department of Commerce. The NIST process model labeled NIST SP 800-101 lays out the digital evidence procedures in four steps: Preservation, Acquisition, Examination and Analysis, and Reporting. These four steps provide the necessary steps for the digital evidence process model as a suggested way to evaluate mobile device information. This is useful, but not complete for law enforcement investigators.

1.2.5 Combined Aquisition and Analysis Methodologies

Hitchcock et al[2] 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 next tier is when 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.

This tiered approach is based on a Computer Forensic Field Triage Process Model proposed by Rogers et al [5] 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[2] 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.

Shaw - One example is to note encrypted compressed files for review later.

Rogers - This process in no way supersedes the ability or need to perform a full forensic examination at a full-featured digital forensic lab.

Ajijola et al, 2014[1] also proposed a new process model that is a hybrid of both models with the resulting combination being much more effective than either of its individual parts.

In the research for combining the NIST and ISO guidelines, Ajijola et al[1] explores the commonality, differences, and limitations of each model. Although both models follow the Auditability, Repeatability, Reproducability, and Justifiability requirements, as well as the Confidentiality, Integrity, and Availability standards, they individually lack some necessary phases to enable them to be used separately. The NIST process model lacks the Identification and Collection phases, while the ISO process model lacks Examination, Analysis, and Reporting aspects of a full Digital Evidence processing model.

The combination of these two approaches, as suggested by Ajijola et al[1], provides a new five step approach: Identification, Collection and Acquisition, Preservation, Examination and Analysis, and Reporting. These steps provide a more comprehensive approach that law enforcement can use to fulfill its evidenciary duties in an investigation. When both process methods are used, the goals approach a full set of tasks from initial on-scene evalutation to the end of the in-lab digital forensics investigation.

1.2.6 Digital Evidence Backlog

IDC forecasts that by 2025 the global datasphere will grow to 163ZB (see Figure 1) (that is a trillion gigabytes). That's ten times the 16.1ZB of data generated in 2016. All this data will unlock unique user experiences and a new world of business opportunities.

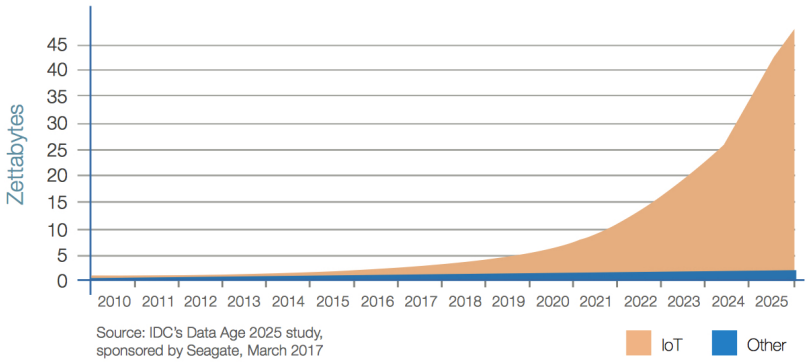

Figure 1: IOT Device Data Growth

Some of the current challenges in digital forensic investigations are directly

related to the amount of data being created. As Lillis et al[3] 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[4] 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[4] 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[4] 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 chal-

lenge, but also new opportunities. Lillis et al[3] 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.

Hitchcock et al[2] - 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.

1.2.7 Digital Evidence Triage

The summary of the research done by Hitchcock et al[2] 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.

One digital evidence triage method proposed by Shaw et al[6] 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[6] extolls the need to reduce digital forensic evidence analysis backlogs, especially with the evolution of big data and the

proliferation of digital devices.

The proposal for a practical and robust method by Shaw et al[6] 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 exists 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[6] 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[6] 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.

Rogers - The CFFTPM was created to enhance the investigators ability to obtain useful information at execution time of a warrant at the suspect's dwelling or work. The process is designed to be used in the first few hours of the investigation, especially during the first suspect interview and search execution phase of the investigation. It is known that suspects are more likely to divulge more information and be more cooperative in that environment (Yeschke 2003[8]). As well, location of and presentation with suspect "triggers" from the potential evidence increase the suspect's willingness to talk and cooperate while on site. ¡NEED CITE¿.

The foci of the CFFTPM are immediately finding usable evidence, identifying victims at acute risk, guiding the on-going investigation, identify potential charges,

and accurately assess the offender's danger to society.

Rogers - The CFFTPM is broken up into phases, two of which have sub-phases (see Figure 2). The main phases consist of Planning, Triage, Usage/User Profiles, Chronology/Timeline, Internet Activity, and Case-Specific Evidence. Usage/User Profiles are broken down into three sub-phases: Home Directory, File Properties, and Registry. These are important and help distinguish user specific activity and permissions. The Internet Activity phase is also broken down into three sub-phases: Browser Artifacts, Email Artifacts, and Instant Messenger Artifacts. These also help establish user activity. Some importance is explicitly stated to skip based on type of investigation and prioritizing the investigation.

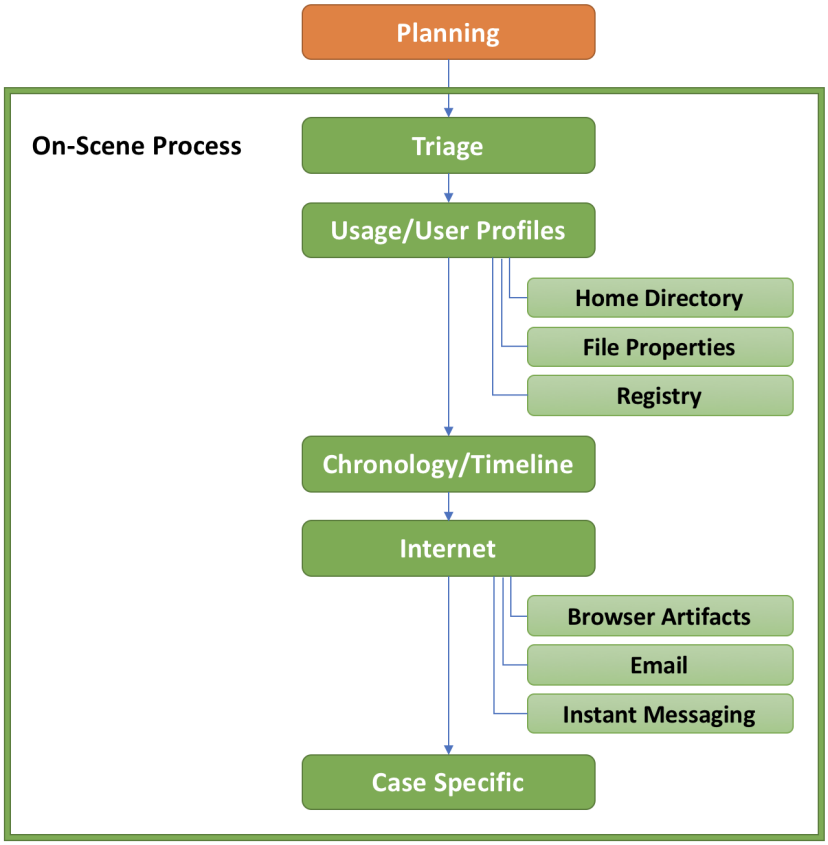

Figure 2: Computer Forensic Field Triage Process Model (CFFTPM)

1.2.8 How SEAKER Can Help

The research of Hitchcock et al[2] should be referenced for a good process starting point for digital forensic labs.

Rogers - This is where the SEAKER portable triage device can help by evaluating every aspect and prevent the on-site investigators from skipping or deprioritizing critical potential evidence.

Rogers - This is also an area where the SEAKER portable triage device can help eliminate some of the potential problems, for instance technical prowess of the on-site investigators and proper lab equipment on-site.

Rogers (CFFTPM) - Some particular aspects of the phases are critical to investigators in revealing evidence or potential evidence for the SEAKER portable traige device. Usage/User Profile information is extremely important. This includes the need to be able to view and search files, folders, registry keys, and file properties associated with a particular user. The Internet Activity artifacts also become very useful, especially in the case of child pornography. The browser, email, and Instant Messaging artifacts can lead directly to potential charges. Finally, a Chronology/Timeline understanding and ability to sort based on it can significantly narrow down the possibilities of which user information and which Internet Activity is the most important and critical to the investigation.

Rogers (CFFTPM) -Implementing this in the SEAKER portable triage device is crucial for simplicity and ease of use. As well, it goes a long way towards having an implementation of SEAKER being understood and adopted. Reporting is also a critical need and is implemented in a way that will enable digital forensic investigators to provide early information to investigators and prosecutors. This helps alleviate the need to wait until the backlog of digital evidence is cleared to get any information from case-specific digital evidence.

2 Background

2.1 Legal Necessity and Implications

Search warrant write blockers chain of evidence

2.2 Technical Necessity and Implications

Raspberry pi Cost human expertise DEFR, DES, Digital Forensic

3 Development of SEAKER Device

great novelty collaboration with industry HTTF

3.1 Conception

Masters Security class SCHTTF project proposal collaborated with SCHTTF

3.2 Setup Script For Raspberry Pi

3.3 Rules For Mounting

mention write blockers from chapter 2

3.4 Code for Searching Device

3.5 Tools Used for Development

4 Experimental Results

4.1 Prototype Demonstration

in-class example at end of semester

4.2 Results

graph of latency estimates of time vs data Gb examples of running on Frank disks

5 Conclusions and Future Work

6 Appendix

6.1 SEAKER Setup

The following set of instructions will detail how to setup the SEAKER environment for the first time. There are three install options that enable SEAKER creators

to prepare the device. See Table 1

1. *Router*: This option is where the Raspberry Pi and the secondary computer are connected directly to the same router, thus allowing the same local DHCP to assign the IPs of both. The secondary computer is used to prepare the micro SD card and to later remotely and securely connect to the Raspberry Pi to complete the setup.

2. *Direct Connect*: This option is where the Raspeberry Pi is connected directly to a monitor and keyboard to enable direct user input via the termnial. The secondary computer is necessary to prepare the micro SD card, but not used to remotely connect to the Raspberry Pi to complete the set up.

3. *Corporate LAN*: This option is almost identical to the *Router* option, but utilizes a corporate network instead of a local router to connect to the Raspberry Pi. This option is the most IT intensive, since the IP address assigned to the Raspberry Pi is often not easily found.

Option 1 (Router)	**Option 2 (Direct Connect)**	**Option 3 (Corporate LAN)**
Hardware Required		
•Raspberry Pi •Mac or Windows Computer •Micro SD card •Router	•Raspberry Pi •Mac or Windows Computer •Micro SD card •Monitor •Keyboard	•Raspberry Pi •Mac or Windows Computer •Micro SD card
Software Required		
•Raspian Stretch Lite •ssh client •Disk Imaging software •IP Scanning software	•Raspian Stretch Lite •ssh client •Disk Imaging software	•Raspian Stretch Lite •ssh client •Disk Imaging software

Table 1: SEAKER set up: Required Hardware and Software

The following creation process is required to setup the SEAKER device to the specifications outlined in this paper. The Figure 3 outlines the general process, while the specific steps are listed below.

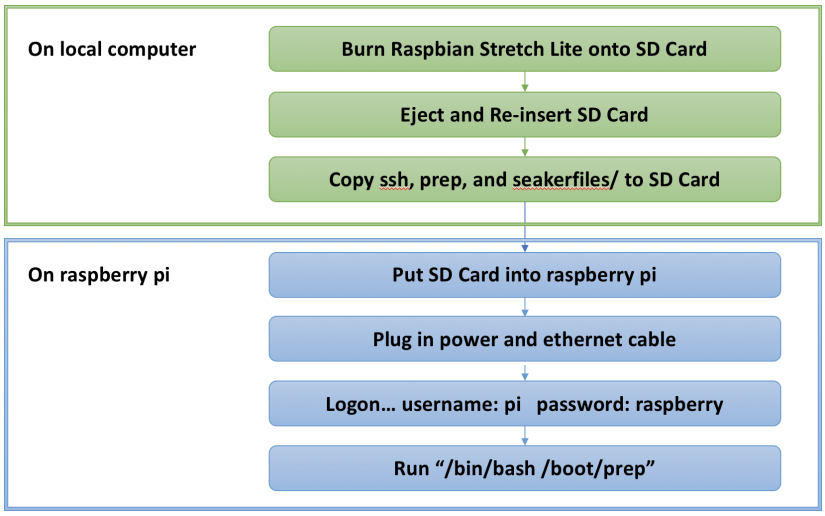

Figure 3: SEAKER Creation Process

1. Download the latest Raspbian Jessie Lite operating system (https://www.raspberrypi.org/downloads/raspbian/). Note the location where file is saved.

2. Download the most recent copy of prep.sh and ssh. These files contain SEAKER code.

   - prep.sh file location: https://s3-us-west-2.amazonaws.com/seaker/prep.sh
   - ssh file location: https://s3-us-west-2.amazonaws.com/seaker/ssh

3. Open prep.sh and edit the default configuration information (shown below). At minimum it is recommended to change the Raspberry Pi and WiFi passwords.

```
# CONFIGURATION SETTINGS
# Raspberry Pi Password
PI_PASSWORD="raspberry"
# WiFi Network Name
WIFI_NAME="SEAKER01"
# WiFi Network Password:
WIFI_PASSWORD="raspberry"
# IP address which is used to access SEAKER web page
WIFI_ROUTER_IP="192.168.101.1"
# DHCP Range (How many connections can be made simultaneously)
WIFI_ROUTER_DHCP_RANGE="192.168.101.50 192.168.101.100
```

   NOTE: It is recommended to change the WIFI Name and IP Address when setting up multiple SEAKER environments over time to ensure each environment has unique identifying information.

   For example: If setting up three SEAKER environments, configuration could be:

   (a) Name: SEAKER01, IP Address: 192.168.101.1

   (b) Name: SEAKER02, IP Address: 192.168.102.1

   (c) Name: SEAKER03, IP Address: 192.168.103.1

4. Insert micro SD card into computer. An adapter will be required.

5. Open disk imaging software. Map to the Raspbian Jessie Lite file location, choose the micro SD card as the destination, and select to burn the image. (Refer to the chosen imaging software documentation for specific instructions on using this tool)

6. Map to the micro SD card in file explorer. Copy ssh and prep.sh onto the micro SD card.

7. Remove the micro SD card from the computer. Insert the card into the Raspberry Pi.

8. Power on the Raspberry Pi by plugging it in with the power cord.

9. Identify the local IP Address of the Raspberry PI:
   If installing with Option 1 (router):

   (a) Plug the Raspberry Pi into the same router being used by the Windows or Mac computer.

   (b) Use the IP Scanning tool on the computer to find the local IP Address of the Raspberry Pi. The Manufacturer should be Raspberry Pi Foundation.

   If installing with Option 2 (Direct Connect):

   (a) Connect the monitor and keyboard to the Raspberry Pi.

   (b) Login using the default username (pi) and password (raspberry).

   (c) Enter the following command to retrieve the local IP Address:

   ```
   ifconfig eth0
   ```

   If installing with Option 3 (Corporate Network):

   (a) The MAC Address of the Raspberry Pi is required. This can be located on the original Raspberry Pi box.

   (b) For Windows systems, open a command prompt and enter the command below. Replace the "c8:26:3b:d2:63:d5" sequence with the MAC Address of the Raspberry Pi being configured. Use the following command:

   ```
   arp -a | findstr "c8:26:3b:d2:63:d5"
   ```

   (c) For Unix or Linux systems such as Apple or Ubuntu, open a terminal window and enter the command below. Replace the "c8:26:3b:d2:63:d5" sequence with the MAC Address of the Raspberry Pi being configured. Use the following command:

   ```
   arp -a | grep "c8:26:3b:d2:63:d5"
   ```

10. If installing with Option 1 or 3:

    - SSH into the Raspberry Pi from the laptop or desktop computer.
      - If using a client such as Putty, enter the local IP address of the Raspberry Pi, choose SSH and connect. Click `OK` or `Yes` on the security warning.
      - If using a command line utility such as Bash enter the following at the prompt:

        ```
        ssh pi@<ip_address> -l pi
        ```

      - Login using the default username (`pi`) and password (`raspberry`).

11. Run the preparation script:

    ```
    /bin/bash /boot/prep.sh
    ```

12. Wait for the Raspberry Pi to finish running the script and rebooting. The Raspberry Pi should now be configured as a SEAKER and be up and running.

6.2 SEAKER Usage

After collecting all the media devices at the scene, the investigator will triage them with SEAKER. Here are the steps:

(1) Connect the RP to the power, and after waiting for about a minute to let it start, connect to the RP's hotspot (WiFi network). Depending on the setup, the WiFi's SSID will be "SEAKER01" or "SEAKER02" etc. (if there were to be more than one SEAKER at the scene). See Figure **??**. Note that the hotspot is password protected; again, the password may be configured.

(2) At the same time the investigator may connect all the media devices to the RP. This may be done concurrently with the previous step. Note that in order to examine a HD it will be removed from the computer, and connected to the RP; this may be done through a write-blocker interface but it is not necessary.

Figure 4: Investigator's handheld view: using a browser, connect to `http://seaker01.local`

(3) Once connected to the hotspot, the investigator will open any web browser on their handheld, and direct it to go to `http://seaker01.local` . We decided to allow access through a web browser as this is the most universal way to connect; any device (iPhone, iPad, Android, laptop, etc.) can connect to a hotspot and open a browser. Once the browser establishes the connection, the user will see

Figure 4. Note that the keywords (or regular expression patterns) present in the "Type in Search Terms:" can be pre-loaded before arriving at the scene, or changed/updated at the scene.

The regular expression can be given using the syntax of the `grep` utility. For example, if we want to find occurrences of either 'two' or 'too', we use `t[wo]o`; if we want to find every word that start with capital letters, we use `^[A-Z]`; if we want to find words where number 9 is the last character of the line, we use `9$`. There are a vast number of possibilities; we can also replace `grep` with `egrep` that has an even richer syntax.

(4) Once any storage devices that are found at a search warrant scene are connected to the RP, the investigator will typically wait for a few minutes (we have seen times up to 10 minutes for 1Tb disks with millions of files) for the file list to be built. The search will then be carried out very quickly: essentially, `grep` browses the file list, line by line, outputting those lines that conform to at least one pattern specified in the "Type in Search Terms:" window. Once this finishes, the investigator will have the results presented as in Figure 5.

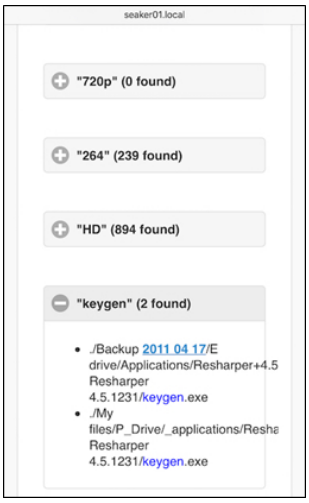

Figure 5: Investigator's handheld view: the results of the search of a particular device.

The filenames themselves can be incriminating evidence, such as in Child Pornography (CP) cases, where the material has a commonly used naming convention, e.g., "lolita" which can be found with the grep pattern `.*lolita.*` ('`.*`' means the following: '`.`' (period) matches any single character of any value, except

a newline, and '*' (asterisk) matches zero or more of the preceding character or expression) or simply `lolita`. This can be used by the investigators to question the suspects. The questioning usually takes place at the same time as the forensic examiners triage the evidence, and one of the requirements of SEAKER was to be fast so that investigators can start getting intelligence quickly from the initial processing of the scene.

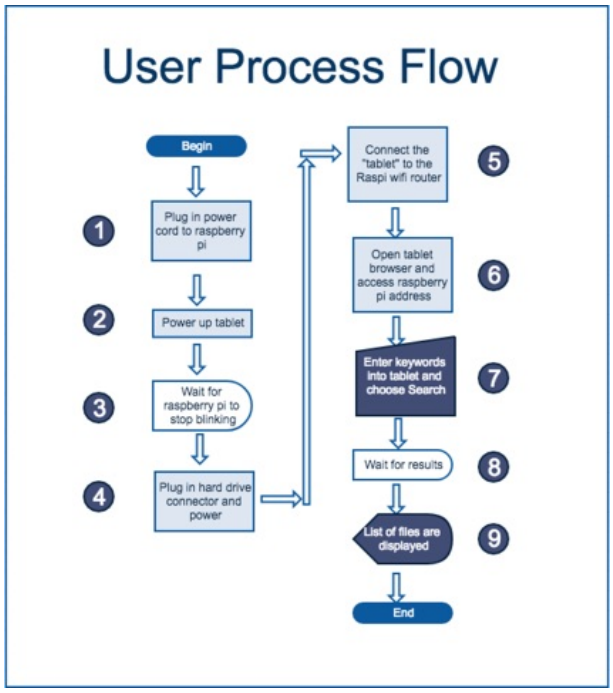

Figure 6: The functionality of SEAKER from the user perspective.

6.3 Code

Some useful code.

References

[1] Akinola Ajijola, Pavol Zavarsky, and Ron Ruhl. A review and comparative evaluation of forensics guidelines of nist sp 800-101 rev. 1: 2014 and iso/iec 27037: 2012. In *Internet Security (WorldCIS), 2014 World Congress on*, pages 66–73. IEEE, 2014.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] J Williams. Acpo good practice guide for digital evidence. *Metropolitan Police Service, Association of chief police officers, GB*, 2012.

[8] Charles L Yeschke. *The art of investigative interviewing: A human approach to testimonial evidence, Second Edition*. Butterworth-Heinemann Boston, 2003.

7 STOP HERE - Abstract

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

8 Meetings with Frank

8.1 May 18th, 2018

My meeting with Frank Lyu, a civilian working for the Ventura County Sherriff'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.

9 Background

9.1 Introduction to Terms

9.1.1 DEFR and DES Investigation Roles

DEFR is Digital Evidence First Responder

DES is Digital Evidence Specialiat

9.1.2 Digital vs. Physical Evidence

I believe there should be a section here that examines (or at least introduces) the differences and similarities between regular physical (non-digital) evidence and digital evidence. Including in the analysis is the metaphore of how physical evidence is handled (bags, DNA, fingerprints) and how that directly relates to the digital evidence model. Contamination must be avoided.

9.1.3 Reactive vs. Proactive Digital Forensic Investigation Processes

Reactive digital forensic investigation processes are utilized after an offense has been committed to help identify the charges and suspects. This is the most common process for digital forensics. The *proactive* digital forensic investigation processes are to attempt to detect before or during an active offense is committed. This is not a job for a typical law enforcement investigator.

This research is based on the reactive digital forensic investigation process in the hopes of reducing the digital forensic lab backlogs across the country and world in two ways. The first way is to reduce the amount of digital evidence acquired for the digital forensics lab by enabling efficient and effective on-site triage to occur by utitizing the commonality of digital evidence collection and analysis into a single step. The SEAKER digital evidence triage tool enables an initial collection of information and subsequent searches by any number of local, on-scene investigators.

These investigators do not need extensive training in digital forensics to utilize it.

The second way is to help reduce the existing backlog by enabling a faster, more streamlined approach to initial potential evidence gathering and reporting. This approach utilizes the SEAKER digital evidence trial tool to perform an initial acquisition and analysis on every exiting case to provide a "first-look" at the information. This also enables a searching mechanism within a few minutes of plugging in digital evidence media to enable digital forensic investigators a quick review of materials. The process will help with prioritization of evidence, a basic analysis and potentially initial evidence in the form of a report that can be provided to investigators and prosecutors.

10 SEAKER Creation

11 Analysis

11.1 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?

11.2 SEAKER Usage Methodologies

11.2.1 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.

11.2.2 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.

11.3 Automated processes during SEAKER evaluation

11.3.1 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"

11.3.2 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

12 Conclusion and future work

12.1 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
- integrate with Microsoft's photoDNA cloud service
- examine local wifi broadcasts
- build an iPad app to simpler, more guided use

13 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
- swap file review
- find encryption Keys
- bulk extractor ?
- thumb strips of movies
- predetermine search criteria (passwords, pw, etc)
- encrypted password entry?
- get file owner, MAC times
- sort based on user
- sort based on access time
- read registry file (make available for search)
- time search i.e. time=lastweek, time=5/5/18-5/15/18
- internet usage timeline
- autosearch/autofilter
- email search

- for drug search... Spreadsheets, documents, databases, internet purchase strives
- for financial search... Spreadsheets, databases, MSMoney, Quicken
- CRC of any acquired files (for later integrity comparison)

14 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
- enhanced previewing - better than triage (full drive)
- Indecent Image of Children
- Image Hash Databases - dbs of digital forensic labs with image hashes
- obviate the need for write blockers
- early look intelligence gathering
- Immediate feedback loop for onsite investigators
- "Suspect's dwelling"
- securing a conviction of the offender
- protecting future victims
- browser artifacts
- onsite == in situ
- adhere to proven forensic principles

15 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[2])
- graph - collection time vs number and size of files
- figure - see figures in Computer Forensic Field Triage Process model (Rogers)
- graph - speed of collecting evidence from same drive based on microSD card used