Abstract - In information and communication technology with the rapid advancements in the world, crimes committed are becoming technically intensive. When crimes committed use digital devices, forensic examiners have to adopt practical frameworks and methods to recover data for analysis which can pose as evidence. Data Generation, Data Warehousing and Data Mining, are the three essential features involved in the investigation process. - The Daubert process used in the admissibility of evidence contains major guidelines applied in assessing forensic procedures, two of which are testing and error rates. The Digital Forensic Science (DFS) community is growing and the error rates for the forensic tools need to be continuously re-evaluated as the technology changes. This becomes more difficult in the case of mobile phone forensics, because they are proprietary. This paper proposes a unique way of generating, storing and analyzing data, retrieved from digital devices which pose as evidence in forensic analysis. A statistical approach is used in validating the reliability of the pre-processed data. And also discusses a database driven approach that could be used to store data about the mobile phone evidence acquisition testing process. This data can then be used to calculate tool error rates, which can be published and used to validate or invalidate the mobile phone acquisition tools.

Keywords - Digital Forensic, Framework, Data Preprocessing, Daubert process

1. INTRODUCTION

The digital world has penetrated every aspect of today’s generation, both in the space of human life and mind, not even sparing the criminal sphere of the world. Forensic science is the application of science to legal process and therefore against crime. It relates the use of science and technology, in the process of investigation and establishment of facts or evidence in the court of law. When crime is aided by or involves the use of digital device(s), the investigation is categorized under digital forensic
or cyber forensic. If the digital device involved is only a computer or digital storage medium, we refer to the investigation as computer forensic. Computer forensic is a branch of forensic science, whose goal is to explain the current state of the digital artifact. The pool of digital devices used by individuals – for work or for entertainment, on a day-to-day basis, includes cellular phones, laptops, personal digital assistants (PDAs), personal computers, wireless phones, wired landlines, broadband/satellite internet connection modems, iPods etc. Each individual today maintains more than one email account, is a member of many communities, virtual groups, takes active part in chat rooms and other networking sites with his/her identity or under an alias, juggles multiple flash drives and other digital storage media. Departments of the Government and Armed Forces, insurance organizations, telephone industries and banks are a few of the sectors which are eager to track, identify and defend themselves against any digital criminal activities.

Digital Forensic Research Workshop (DFRWS) has defined Digital Forensic Science as “the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations”

Digital Forensic Science covers computer forensics, disk forensics, network forensics, firewall forensics, device forensics, database forensics, mobile device forensics, software forensics, live systems forensics etc. Digital Forensic has been described as incident(s) specific and practitioner driven advances which are developed and then applied. The DFRW has identified media analysis as one of the three main distinct types of digital forensic analysis, the other two being Code Analysis and Network Analysis. This paper introduces a framework for the digital forensic investigation process of physical storage device. It also takes a specific case of accessing the flash drive as a device and analyzing its contents. The paper details the preprocessing steps adopted to bring out information of the data stored on the flash drive.

2. FRAMEWORK FOR DIGITAL FORENSIC INVESTIGATION PROCESS OF PHYSICAL STORAGE DEVICES

A framework, for seamless communication, between the technical members of the digital forensic investigation team and the non-technical members of the judicial team, is very necessary. Defining a generic model for digital forensic investigation, sometimes pose a problem taking into account the varied devices available today. This framework is logical in its outline, scientific in its approach though it is to be adapted to comply with all the legal requirements of the country where the incident has occurred. It charts to add value in the specific case of portable storage digital devices. Made up of six stages, it is practical in approach, easy to implement when the digital device involved is any portable, storage device.

Stage 1:
Preparation: The main focus is acknowledging then role of digital storage device(s) in the identified or untoward incident. This step recognizes the presence or absence of the digital forensic investigation. All suspected physical storage devices are to be physically secured to prevent tampering. The concerned authorities are to be notified about the presence of possible evidence(s) and the need for examination of the same, and hence permission to access the device. In case the evidence needs to be removed from the premises or site of the activity, steps for obtaining the necessary permissions for the removal are to be identified and executed. On the whole, based on the nature of the incident or crime, the investigation steps are to be chalked out.
Stage 2:
Collection and preservation of digital device: The device collection phase opens with the identification of the ownership of the device along with the identification of supposed users of the device. All the digital devices and any other supporting evidences about the usage of these devices that are present at the scene of crime are to be confiscated for data collection. In case the physical device is password protected, the software necessary for accessing the device contents is identified and verifying that it does maintain the integrity of the data as it works on accessing the device. The device contents should be duplicated or imaged maintaining the integrity of the data in the device. Each step of the activity should be documented.

Stage 3:
Data extraction and preprocessing: The device/disk that has been imaged or duplicated is to be accessed and examined for the presence of any hidden or encrypted data and system related data. Required software tools are to be used to decrypt or access the data. These tools should not tamper the original data. Ensure that nothing will/shall be written on to the device that is under scrutiny. Based on the nature of the incident, the investigation is to be categorized as goal based or non-goal based. The data should be extracted from the digital device and the steps for the preprocessing the data is to be outlined, justifying the reason for the same. The software required for the process is to be identified. All through the stages, concern about maintaining the integrity of data should be the key focus and each step is to be validated before executing. Documentation of the activities carried out should be precise and justified as this would act as the base document for justifying the integrity of the presence or absence of evidence leading to the crime.

Stage 4:
Data examination and analyses: Before the data is subjected to examination and analyses, the data is to be crosschecked for authentication and integrity. The analyses that can be carried out on the extracted data, based on the nature of the data, are to be considered along with the required tools to perform the same. On justifying the analysis methodology, the actual analysis is to be carried out until stable results are achieved. Interpretation of data is the most difficult step, while at the same time the most important step in this flow.

Stage 5:
Reporting and documentation: Though this has been cited as the stage 5, it is a continuous process, which needs to be reviewed and updated finally, before presentation in the court of law, for completeness and accuracy. The validity and the acceptance of the process or methodology in the scientific community should also be explored. Documentation of the analyses, conclusions and assumptions if any, are also of importance. The limitations of the procedures/analysis carried out are to be outlined clearly.

Stage 6:
Presentation in the court of law: The main focus of this step is to prove the presence or absence of digital evidence, from the digital devices collected from the scene of the incident under examination, in the court of law. While computer forensics is highly technology specific, people handling law in the court of justice are not technology specialists. Hence it is very important for technology specialists to understand the ramifications of the legal world and at the same time, communicate effectively and clearly the complete digital investigation process, emphasizing on the analysis of the findings. The documentation of the entire may also be submitted in the court of law to cross-examine the steps adopted during the investigation process. While this may suffice the needs of the court to arrive at a decision, it may sometimes be required to complete further analysis or redo a phase, as required by the court, to support any issues.
3. MOBILE PHONE USAGE

Mobile phones are widely used in the United States. In the first six months of 2006, the Cellular Telecommunication and Internet Association (CTIA) stated that there were 219.4 million U.S. wireless subscribers, and wireless communication has penetrated more than 72% of the total U.S. population. CTIA also explained that customers used 857 billion Minutes of Use (MOUs). Additionally, CTIA reported that 64.8 billion SMS messages were sent, an increase of 98.8% from the first six months of 2005 (“CTIA Quick Facts.” 2006). Digital evidence is becoming important, where 80% of current court cases have some sort of digital evidence associated with them (Rogers, 2006, p.1). Summers (2003) explained “In the past five years, dozens of murderers have been convicted partly as a result of evidence about their mobile phones or those of their victims”. Mobile phones are becoming more than just simple phone devices. Numerous technologies are being integrated within them such as Bluetooth, digital cameras, Infrared, General Packet Radio Service (GPRS), E-mail and more.

Evidence needs to be acquired from mobile phones when needed in a forensically sound manner. In the realm of digital forensics, software tools have dominated the market in the acquisition of digital evidence from mobile phones. These tools have not been tested and have no published error rates. The only notable tool testing initiative for mobile phone forensics was performed by the National Institute of Science and Technology (NIST). This initiative is by no means complete, especially since they were only able to test a limited number of mobile phones, seventeen to be exact (Jansen, Wayne, Cilleros & Daniellou, 2005).

4. FORENSIC TOOL TESTING

Tool testing programs have been taken into consideration by various organizations. Tool testing is important from an Information Technology (IT) perspective to make sure that software and hardware operate as expected. The Institute of Electrical and Electronics Engineers (IEEE) has established standards since 1993 for tool testing. The International Organization for Standardization and the Electro technical Commission (ISO/IEC) then established the General Requirements for the Competence of Testing and Calibration Laboratories (ISO/IEC 17025) in 1999 (“General Testing Methodology.”, 2001).

NIST’s Computer Forensics Tool Testing (CFTT) program had the right intentions from a technical perspective, and NIST’s (“General Testing Methodology.”, 2001) states that the general requirements to test a tool are:

1. Establish categories of forensic requirements
2. Identify requirements for a specific category
3. Develop test assertions based on requirements
4. Develop test code for assertions
5. Identify relevant test cases
6. Develop testing procedures and method
7. Report test results

One can apply the aforementioned list to test a tool that is designed to work for a single, specific purpose, in an environment that is absolutely constant. However, in the case of mobile phones, numerous variables such as phone model, phone provider, cables used, even the fact that the mobile phone is on (data on a cell phone continuously changes when it is turned on) are all important factors that need to be properly documented.

5. TESTING IN DNA FORENSICS

Forensic sciences have flaws. DNA forensics for example, has been widely accepted, yet even the results obtained from DNA forensics are not perfect. In a study by Saks & Koehler (2005), the following factors were shown to play a role in the wrongful conviction of 86 DNA exoneration cases:

1. 71% Eyewitness error
2. 63% Forensic science testing errors
3. 44% Police misconduct
4. 28% Prosecutorial misconduct
5. 27% False/misleading testimony by forensic scientists
6. 19% Dishonest informants
7. 19% Incompetent defense representation
8. 17% False testimony by lay witness
9. 17% False confessions

The two interesting statistics noted above are numbers 2 and 5. One can only imagine what the statistics would be like in the case of DFS, and it is our duty as Digital Forensic scientists to decrease those testing errors.

6. SYSTEMATIC DATABASE DRIVEN TESTING METHODOLOGY

The authors of this paper created a systematic database driven testing methodology for mobile phone tool testing. This will contribute to establishing repeatability estimates of the various tools that are used when acquiring digital evidence from mobile phones. With that comes a number of issues, mainly that mobile phones are proprietary. Therefore, a robust testing methodology for all mobile phones should take that into consideration.

Cellular Phones are Proprietary

New mobile phone models are released frequently by various corporations. It would be a difficult task to keep up with all these phone models and their various proprietary features. This poses a challenge in mobile phone forensic tool testing because a robust evidence acquisition system should be able to forensically acquire evidence from all mobile phone models even with the phone’s software and hardware proprietary natures.

Some of the proprietary mobile phone characteristics are outlined below. When dealing with mobile phone forensics, the following are important factors that should be recognized when performing a forensic acquisition:

1. Mobile phones have proprietary file systems.
2. Mobile phones have proprietary file transfer protocols.
3. Mobile phone providers lock down certain features of the device.
4. Different mobile phone providers might install different operating systems on the mobile phone device.
5. Cables used in the forensic acquisition of a mobile phone can be different.
6. The mobile phone device’s clock changes data continuously on a device.
7. Different mobile phones have different features.
8. A mobile phone being used is being provided a service through a carrier, and there are numerous carriers.
9. Applications can be installed on certain cellular phone models.

Process model for Cellular Phone Tool Testing

Based on the tool testing literature, a simplistic tool testing process model was developed. We envision the implementation of this process model as a programmatic database driven system. This process model is delineated in Figure 1.

As shown in Figure 1, the process model is simplistic in nature. The data that needs to be wrapped around that process model can be tedious. The above scenario is not necessarily new, as it adheres to the NIST tool testing methodology. The model simply takes the usual tool testing standards and tweaks them so that the process model is programmatically driven by a database system. Based on the process model and the proprietary nature of mobile phones, a relational database schema was developed to aid in illustrating the different data requirements for the forensic tool testing of mobile phones. Entity Relationship Diagrams (ERDs) are useful in representing data requirements. They adopt a more natural view that the real world consists of entities and relationships (Chen, 1976).
Roshan Singh Thakur, Khyati Chourasia and Bhupendra Singh Thakur

The formulated database Entity Relationship Diagram is shown in Figure 2.

Figure 2 - Proposed ERD
**ERD Explanation**

The ERD is merely the representation of the data that should be stored in the database. The ERD above enables the users of a system to have access to the following data (assuming that the data has been entered into the database):

- The various mobile phone models
- The various features supported by each specific mobile phone
- The various test case scenarios
- The various mobile phone forensic acquisition tools
- The various features supported by each cellular phone acquisition tool
- Various data cables used in the forensic acquisition process
- The mobile phone models supported by each data cable
- The locked down features of a mobile phone that might be locked down by a carrier
- The various mobile phone carriers
- The operating system of a mobile phone installed by carrier for a specific phone model
- The various applications that might be installed by carrier for a specific mobile phone model
- Whether or not the overall test case passed or failed
- The features of the mobile phone tool test that passed or failed (For example, SMS acquisition passed, but call history failed)
- Various applications that could be installed on a mobile phone

The above data can help in obtaining specific information about the mobile phone being tested. It can also help in either validating or invalidating a forensic tool. For example, if a specific test case scenario were repeated, with the same conditions, yet with varying results, the tool would be deemed invalid. To know whether or not a tool is valid or invalid, error rates for the tools will have to be calculated.

**Error Rate Types**

Tool errors have been discussed in the literature (Carrier, 2002, 2003). The focus of this paper is not to create a topology of errors, but to illustrate a practical approach for the calculation of testing error rates. Using the database driven approach, one would be able to calculate the General Error Rate (GER) and the Feature Error Rate (FER) for every mobile phone test case scenario. These error rates can be calculated for every test case scenario as follows:

\[
\text{General Error Rate (GER)} = \frac{\text{# Unsuccessful forensic acquisition processes}}{\text{# of Test Cases}}
\]

\[
\text{Feature Error Rate (FER)} = \frac{\text{# Unsuccessful feature acquisition processes}}{\text{# of Feature acquisition processes}}
\]

As a motivating example, assume that a mobile phone was analyzed and a test case scenario was performed. Different mobile phone devices that are of the same model were then used with the same test case scenario. For example, if 10 different mobile phones that are all Sony Ericsson K800i’s were analyzed using the same test case scenario and the acquisition system worked only 4 times, then the GER = 6/10, which would equal 0.6 or 60%. Furthermore, if the software only fully worked 4 times, but it was able to acquire Short Message Service (SMS) messages, 7 times, then the FERSMS= 3/10 which would be 0.3 or 30%.

7. DISCUSSION OF USING A DATABASE DRIVEN APPROACH

The proposed process model and database schema are advantageous for numerous reasons. Primarily, we as DFS scientists do not have access to any information about the tool error rates. This violates Daubert’s second procedure as outlined by Carrier (2002). A database solution that logs all of the forensic examinations and tool testing procedures for the various mobile phones can help in the establishment of some type of calculated error rates based on the historical stability of these tool sets. This can also help in establishing a reliability measure for the various tool sets. Once Computer Forensics professionals become aware of the error rates of the forensic tools, it will help them become more...
objective with their decisions as to what tools perform the best functions. Furthermore, as mobile phone cases become more prominent in the courts, expert witnesses will be called upon. As an expert witness, it is vital to recognize the various error rates of the forensic tools to ensure that decisions are being made beyond a reasonable doubt.

Furthermore, having a standardized database driven approach for mobile phone forensics tool testing can help document the various testing scenarios that have been performed. The documentation of all the tests and their results create a historical repository that can be used for trend analysis, such as data mining. Statistical techniques can be applied to these historical results to help formulate predictions about the future forensic analysis of mobile phones by type, model, carrier etc. These results can further be used to ascertain the benefits inadequacies of the various mobile phone Forensics tools.

Proposed Model Challenges

There are certainly some issues with using our proposed model. The first question that can be posed is how often are forensic professionals going to forensically acquire data from a mobile phone, and record the results using our anticipated database driven methodology? The authors believe there are two answers to this question. Primarily, we notice an increase of mobile phone ownership amongst college students. Therefore, scientists have access to a significant number of test cases at their disposal in learning institutions. Furthermore, the number of mobile phone models, and the number of mobile phones to be analyzed can come from that random sample of students through a research initiative. A database solution as such can be used by various investigators for either looking up past test results, or for adding their test case results. Of course, for that to occur, these investigators should be willing and able to provide us with that data. The solution the authors propose is not a definitive resolution to the error rate problem, but should spark some interest for further research in the area. Again, our purpose is to be able to find ways of attaining tool error rates, and furthermore, an estimate for the repeatability of the results that each mobile phone forensic acquisition tool provides. Another issue with our proposed model is keeping the database up to date with the latest versions of the forensic acquisition tools and their capabilities. Some vendors release updates on a very frequent basis and their cooperation might be needed to keep the database current. A possible solution to this problem is to get the vendors involved in the process to continuously update the information in the database.

8. CONCLUSION

Mobile phone forensics is a novel field. When analyzing a mobile phone for forensic evidence, the process of doing so is different than the traditional computer forensics model. As forensic scientists, we should always be aware of the laws that deal with the admissibility of evidence, mainly the Daubert guidelines outlined by Carrier (2002) and in Daubert v Merrell Dow Pharmaceutical (1993). The first and second Daubert guidelines that deal with tool testing and error rates are the two major issues that this paper focused on. A database driven approach for the documentation of the mobile phone forensics procedures can ameliorate the process of documenting the testing methods. This will assist in acquiring results on the various test cases. These results can promote the calculation of tool testing error rates. This information will help DFS scientists validate or invalidate mobile phone forensics evidence acquisition tools and help expert witnesses make better decisions, beyond reasonable doubt about the evidence acquired from the mobile phones.

REFERENCES


